n.º 11 | dezembro 2013

Saber & Fazer

Telecomunicações

design thinking, usabilidade, processo criativo, empatia, prototipagem, *user centered design, computer vision*, realidade aumentada, mobilidade, aplicações, meo, Move4Health, Online-Gym, MarvIN, consistência atómica, consistência eventual, latência, *throughput*, alta-disponibiidade, fiabilidade, *cloud computing, service broker, platform-as-a-service,* APIs, interoperabilidade, WebRTC, HTML5, VoIP, configuração de sistemas, conformidade, política de segurança, perfis, *standards*, *baselines, sticky policies*, sistema criptográfico, IBE, controlo de acessos, RBAC, linguagem de políticas, auditoria de ficheiros, identificação, autenticação, segurança, deteção de intrusões, biometrias comportamentais, *keystroke dynamics*, *mouse dynamics,* privacidade, M2M, *ambient intelligence,* Segura, SIGO®, controlo de acessos, operacionalização de recursos, segurança de edifícios, pagamentos móveis, NFC, segurança UICC, IMS, SBC, interligação, automatização de testes, SIP, 3GPP, LTE, PCC, PCRF, PCEF, TDF, CAC, RAN, eNB, EPC, Wi-Fi, ip-Raft, 3GPP, WLAN, LTE, EPC, *software defined networking, OpenFlow,* FTTx, topologias de rede de acesso, infraestruturas de rede, materiais de rede passivos, projeto de rede, FTTx, infraestruturas de rede, CAPEX, OPEX, *pay-as-you-grow, future-proof, unbundling*, rede de acesso

PT INOVAÇÃO

Saber & Fazer

Telecomunicações

design thinking, usabilidade, processo criativo, empatia, prototipagem, *user centered design, computer vision*, realidade aumentada, mobilidade, aplicações, meo, Move4Health, Online-Gym, MarvIN, consistência atómica, consistência eventual, latência, *throughput*, alta-disponibiidade, fiabilidade, *cloud computing, service broker, platform-as-a-service,* APIs, interoperabilidade, WebRTC, HTML5, VoIP, configuração de sistemas, conformidade, política de segurança, perfis, *standards*, *baselines, sticky policies*, sistema criptográfico, IBE, controlo de acessos, RBAC, linguagem de políticas, auditoria de ficheiros, identificação, autenticação, segurança, deteção de intrusões, biometrias comportamentais, *keystroke dynamics*, *mouse dynamics,* privacidade, M2M, *ambient intelligence,* Segura, SIGO®, controlo de acessos, operacionalização de recursos, segurança de edifícios, pagamentos móveis, NFC, segurança UICC, IMS, SBC, interligação, automatização de testes, SIP, 3GPP, LTE, PCC, PCRF, PCEF, TDF, CAC, RAN, eNB, EPC, Wi-Fi, ip-Raft, 3GPP, WLAN, LTE, EPC, *software defined networking, OpenFlow,* FTTx, topologias de rede de acesso, infraestruturas de rede, materiais de rede passivos, projeto de rede, FTTx, infraestruturas de rede, CAPEX, OPEX, *pay-as-you-grow, future-proof, unbundling*, rede de acesso

PT INOVAÇÃO

Título
Saber & Fazer Telecomunicações
Revista técnica da PT Inovação

Edição
PT Inovação 2013

Design
Patrícia Gaspar,
Designer de Comunicação

Impressão
FIG – Indústrias Gráficas, SA

Tiragem
750 exemplares
ISSN 1645-8710
Depósito Legal 251344/06

ALCINO LAVRADOR
Administrador Delegado da PT Inovação

A banalização do acesso a banda larga, nomeadamente em situações de mobilidade, e a explosão da internet estão a moldar um mundo novo, com novos atores e novos modelos de negócio, cujos impactos nos operadores de telecomunicações os tornarão irreconhecíveis num futuro próximo.

Pelo lado da procura, os novos consumidores, individuais ou empresariais, mais exigentes e sofisticados, apresentam mudanças de comportamento e requisitos que exigem agilidade e flexibilidade do lado da oferta, mas sobretudo também, antecipação de necessidades.

Os serviços tradicionais *core* perdem continuamente relevância na estrutura de receitas dos operadores, sendo determinante encontrar um posicionamento adequado, capturando o valor proveniente de novos espaços de oportunidade que se criam pela transformação e evolução tecnológica e pela convergência IT/Telecom e, assim, transformar as ameaças em oportunidades.

Os serviços de comunicação, transformados em aplicações, devem ser desacoplados do acesso, permitindo a sua utilização em múltiplos contextos, combinados de diferentes formas e suportados em diferentes plataformas. Neste novo mercado, altamente competitivo, à oferta de preços baixos, equipamentos terminais apelativos e serviços inovadores, deve acrescentar-se fiabilidade, disponibilidade e privacidade.

Com o orçamento disponível dos consumidores para comunicações a ter que ser cada vez mais partilhado com outros *players*, interessa por um lado ocupar o máximo possível das novas cadeias de valor e, por outro, encontrar mecanismos para ganhar eficiência operacional. Fazer os investimentos certos, utilizando racional e inteligentemente os recursos humanos e financeiros disponíveis e cada vez mais escassos, será crucial para desenvolver as novas propostas de valor que possam permitir ganhar a preferência dos clientes e consumidores.

A experimentação, a antecipação de riscos tecnológicos e a partilha por toda a organização do conhecimento adquirido, fazem parte da missão da PT Inovação, como contributo para a antecipação do futuro. Este número de Saber & Fazer contém diversos artigos que endereçam múltiplos domínios e são resultado da atividade de Investigação e Desenvolvimento em projetos de Inovação Exploratória e provas de conceito, procurando identificar as principais características e tecnologias que podem suportar uma oferta diferenciada de serviços e com alto valor acrescentado. Contém também casos de estudo que interessa partilhar para que a partir dos resultados obtidos se possam tirar ensinamentos.

A todos aqueles que tornaram possível a edição de mais este número, clientes e parceiros, com os quais aprendemos continuamente e, sobretudo, aos autores dos artigos, o meu agradecido reconhecimento.

MARCELINO POUSA

Caros Colegas:

É com agrado renovado que escrevo este edital para a edição número 11 da revista Saber & Fazer Telecomunicações, isto porque passámos a barreira da edição 10 com a mesma vontade de partilhar o nosso trabalho e algum do conhecimento que vamos adquirindo com um conjunto de temas que mostram a nossa capacidade para acompanhar o estado da arte nas mais diversas áreas tecnológicas relevantes para o nosso negócio.

Neste número apresentamos uma seleção de temas que esperamos sejam do vosso agrado e que neles possam encontrar ajuda para melhor perceber a nossa envolvente, ou simplesmente motivação para estudar o assunto. Consideramos que se isso acontecer a revista cumpre o seu papel e será um motivo para que continuemos a trabalhar para manter a sua edição em 2014.

A revista está estruturada em 4 secções que, sem pretenderem delimitar áreas tecnológicas, agregam temas próximos. A **primeira secção** contém temas mais teóricos: *Design Thinking*, uma metodologia de criatividade e inovação muito utilizada na criação de produtos e serviços; Visão por computador, técnicas de reconhecimento com uma importância relevante no desenvolvimento dos robots, e novas funcionalidades de interação baseada no processamento digital de imagem para apoio à operação humana em ambientes complexos; Bases de dados NoSQL, taxonomia e *tradeoffs* de desempenho, disponibilidade e consistência de dados; Cloud PaaS, conceito requisitos e limitações; WebRTC, será que esta tecnologia vai ter mais relevância do que a Voz sobre IP para os operadores de telecomunicações?

Na **secção dois** abordamos aspetos relacionados com a segurança da informação, Automatização da verificação de políticas de segurança e deteção de vulnerabilidades; Reforço da Privacidade Através do Controlo da Pegada Digital, artigo que apresenta mecanismos que tendem a diminuir a assimetria de poder e conhecimento entre os utilizadores e os *"service providers"*, na garantida da confidencialidade da informação; Identificação biométrica e comportamental de utilizadores em cenários de intrusão, o artigo propõe novas técnicas de reforço de autenticação de utilizadores; Privacidade, o *M2M killer-issue*? Este artigo aborda aspetos atuais como a privacidade, em cenários M2M, em que a quantidade de informação que um cidadão partilha vai crescer exponencialmente.

Na **secção três** apresentamos aplicações desenvolvidas na PT Inovação, como o projeto Segura, Uma gestão integrada das equipas de operação e da segurança de edifícios; Mobipag, um ecossistema de soluções e serviços de valor acrescentado sobre pagamentos móveis; Solução de *VoIP Peering* em Redes IMS, fundamental para garantir conetividade entre operadores sem recorrer às redes PSTN existentes; finalmente um artigo que apresenta uma solução de Automatização de testes SIP em redes IMS, fundamental para a operacionalização de serviço.

Na **quarta secção**, apresentamos artigos relacionados com as infraestruturas de rede e controlo de tráfego: PCRF e Controlo de Congestão, apresenta novas técnicas de controlo e gestão de trafego de forma a evitar constrangimentos nas redes *wireless*; Integração *Wireless* LAN no core EPC, este artigo aborda as novas funcionalidades discutidas no

3GPP que permitem ao equipamento terminal, de forma transparente, selecionar e mudar de rede de acesso, comutando entre WLAN e redes 3GPP, dessa forma ligando-se à rede de acesso mais vantajosa; *Software Defined Networking* – Prova de Conceito, neste artigo aborda-se o conceito de SDN, o teste da tecnologia *openflow* e qual o seu potencial interesse para a PT. Para finalizar temos dois artigos sobre arquitetura e projeto de redes FTTH, com foco na arquitetura da rede da OI, e a otimização do CAPEX em redes FTTX.

Resta-nos agradecer aos autores que se disponibilizaram a partilhar connosco o seu trabalho e conhecimento, à equipa de revisão e edição, onde destaco a Raquel Nogueira, Inês Oliveira, Clara Guerra, Nuno Seixas e Ricardo Afonso pelo trabalho de suporte e revisão de mais este número. Aos leitores desejamos que encontrem na revista algumas respostas às suas inquietações e alguns sinais que despertem a curiosidade e permitam aprofundar áreas de interesse para a empresa.

01 Design Thinking



INÊS OLIVEIRA LÚCIA MOREIRA NUNO SEIXAS

PALAVRAS CHAVE
Design Thinking, Usabilidade, processo criativo, empatia, prototipagem, *user centered design*

Design thinking é hoje uma das expressões mais em voga nos domínios do conhecimento relacionados com a Inovação e Criatividade. No entanto, o processo de transição da expressão para a implementação implica perceber os seus princípios e conceitos de base, para de seguida se conseguir entender e implementar as diversas técnicas associadas.

Esta metodologia advoga que o processo criativo e, de uma forma mais geral, a resolução de problemas, se foque primeiro no problema a resolver, na identificação das necessidades que se pretendem endereçar. Para que se adquira este tão importante conhecimento é necessário que se estabeleçam relações de empatia com os utilizadores. Um segundo aspeto muito importante é a abordagem iterativa, exposta neste artigo, de execução em pequenos ciclos compostos pelos 5 passos: *empathyze, define, ideate, prototype* e *test*.

Neste artigo irão explicitar-se os conceitos e funcionamento do design thinking, mostrar exemplos de técnicas que podem ser aplicadas para a sua implementação e ainda referir alguns resultados já obtidos com a sua aplicação no contexto da PT Inovação.

Mas a aplicação desta metodologia não está limitada ao desenvolvimento de produtos tecnológicos, estando cada vez mais a ser usada em áreas como a definição de políticas ou remodelação de serviços, pelo que se procura igualmente mostrar a razão desta utilização "alternativa" e alguns exemplos fora do mundo tecnológico.

No final estabelecer-se-á a ponte com a inovação, referindo a importância.

1. INTRODUÇÃO

O design é o processo através do qual um artefacto é criado. Faz parte do dia-a-dia de cada um de nós, daquilo que é, que sentimos e tocamos. O design thinking, em oposição, foca-se no não tangível, no problema, na necessidade. É uma forma de pensar que combina o pensamento criativo e analítico e o aplica na resolução de um problema específico. Esta forma de pensar é organizada numa metodologia que incorpora o espetro de atividades de inovação com o caráter do design centrado no utilizador [3]. A inovação é alimentada por um conhecimento meticuloso, através de observação direta, do que as pessoas querem e precisam nas suas vidas e o que gostam ou não na forma como determinados produtos estão feitos, empacotados, comercializados, vendidos e suportados [3]. Baseia-se na capacidade de considerar em simultâneo: (i) necessidades humanas e novas perspetivas, (ii) material disponível e recursos tecnológicos e (iii) constrangimentos e oportunidades para um projeto ou negócio.

O design thinking é uma metodologia cada vez mais difundida no mundo corporativo, contribuindo para o desenvolvimento da capacidade criativa orientada especificamente às áreas de trabalho em causa (produtos ou serviços) e com uma forte base numa análise qualitativa, bem como para o desenvolvimento do trabalho em equipa.

O processo que rege a metodologia de design thinking divide-se em cinco passos iterativos (não lineares). São eles: *empathyze, define, ideate, prototype e test*.

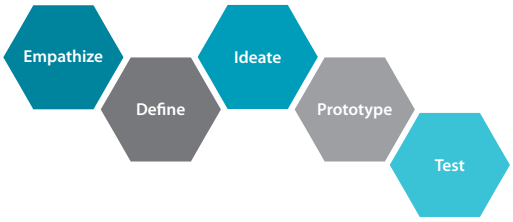


Figura 1. Passos iterativos do design thinking

Na fase de criação de empatia (*empathyze*) devem colocar-se todas as suposições e ideias iniciais de parte e deixar os utilizadores constituírem a fonte de inspiração. Considera-se fundamental mergulhar na experiência do utilizador, observar, questionar e partilhar experiências.

A proximidade com o utilizador permite recolher informação que deve ser aproveitada numa fase de divergência, em que todas as ideias contam, importando a quantidade em detrimento da qualidade. Nesta fase, importa então sintetizar as aprendizagens em necessidades e criar um ponto de vista (*define*), gerar ideias, criar rascunhos da solução, trocar feedback e refletir (*ideate*). Numa fase final de convergência elaboram-se escolhas através de criação de protótipos físicos, experiências e observação (*prototype, test*).

2. METODOLOGIAS ASSOCIADAS

A operacionalização do design thinking prevê a execução de um conjunto de técnicas associadas às várias fases do processo. Em função das características de cada projeto, nomeadamente no que concerne à sua dimensão temporal, financeira e também dos objetivos que lhe estão associados, devem ser selecionadas as técnicas que melhor se adequam a esse contexto.

>>>

Esta metodologia confere um grau de importância central ao utilizador final e ao contexto em que este se insere, facto que tem implicações no desenho e conceção de toda a experiência de utilização do produto/serviço que se está a desenvolver. Assim, o ciclo de aplicação das ferramentas e técnicas nesta metodologia respeita uma orientação aos modelos de *user centered design*, os quais vêm romper com o paradigma clássico de desenvolvimento de *software*, que tende a privilegiar a especificação de requisitos em função da tecnologia e não em função do utilizador final.

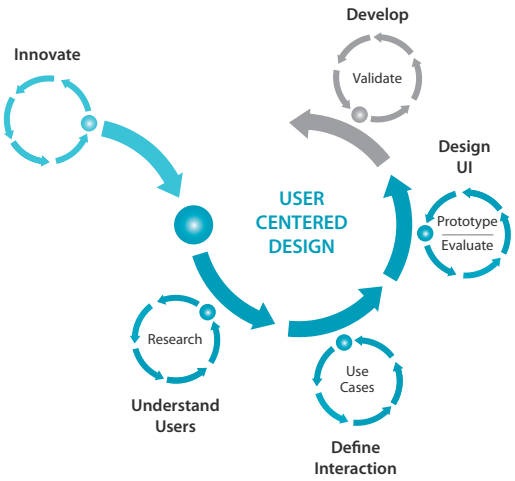


Figura 2. Modelo de *User Centered Design*

User centered design é uma abordagem holística de desenvolvimento de produtos e serviços, que coloca o utilizador no centro do processo e que integra informação para a especificação de toda a envolvente, contexto e características das pessoas que vão efetivamente utilizar o produto ou serviço. [6]

Na primeira fase, são conduzidas pesquisas de carácter preliminar e exploratório, as quais permitem, por um lado obter um enquadramento do tema em análise e por outro, analisar e caracterizar de forma mais detalhada a problemática, expectativas e os perfis de utilizador.

O enquadramento revela-se como fundamental, pois permite à equipa conhecer a visão de vários agentes sobre o ecossistema que se apresenta como objeto de análise e, sempre que necessário, completar esse olhar com novas perspetivas, que permitem por exemplo, o reposicionamento de um produto ou serviço.

As técnicas mais aplicadas para a coleta desta informação são: entrevistas, cadernos de sensibilização e sessão generativa.

A pesquisa exploratória completa este ciclo ao ampliar a explicitação do conhecimento em termos de arquitetura de informação e de organização semântica de conceitos. Poderá ser executada em ciclos iterativos de melhoria, na medida em que a equipa recolhe informação, analisa, sintetiza e identifica se é necessário repetir o ciclo ou integrar novas ferramentas para recolher mais dados e tornar o corpus da pesquisa mais sólido, permitindo iniciar uma nova fase do processo.

As ferramentas mais aplicadas para a coleta desta informação são: personas, blueprints, cartões de insight, mapa de conceitos, mapa de empatia e diagrama de afinidades. Em grande parte dos casos, estas duas pesquisas fundem-se e as técnicas e ferramentas são ajustadas para um passo metodológico só, predominantemente qualitativo.

É também habitual efetuar uma triangulação dos dados obtidos nestes momentos, com vista a identificar padrões e oportunidades, que funcionam como catalisadores para encontrar soluções, e que permitem avançar para a fase seguinte no processo.

A fase de geração de ideias (*ideate*) é o primeiro momento em que a equipa se encontra em harmonia sobre as referências e expectativas que tem para a construção coletiva desse produto ou serviço explorado no projeto. São utilizadas técnicas como o brainstorming, *workshop* e matriz de posicionamento. Um dos resultados mais relevantes desta fase é a consolidação das ideias e grupos funcionais que devem ser prototipados. Para o efeito, é elaborada uma matriz que posiciona as funcionalidades/ideias ao nível de retorno face ao grau de inovação. A aplicação desta matriz é, em alguns casos, adaptada e/ou complementada com critérios específicos do projeto em análise e que podem fornecer prioridades distintas para ciclo de desenho da solução final, como por exemplo funcionalidades que trazem um grau de inovação reduzido, mas que oferecem grande conforto ao utilizador final em termos de experiência de utilização.

A partir desta matriz de posicionamento, a equipa avança para a elaboração de propostas de desenho da solução. As ferramentas mais utilizadas para suportar o desenho da solução final são: prototipagem em papel, prototipagem digital (alta fidelidade funcional e não funcional), *card sorting*, elaboração de *storyboards* e, no caso de se tratar de um projeto de serviços ou de *software* que integre serviços, podem ser também desenvolvidas experiências em cenários físicos para os utilizadores finais experimentarem e darem o seu contributo para o

alinhamento funcional da solução que se está a desenvolver. A este ciclo de prototipagem agrega-se uma etapa de validação iterativa, a qual prevê a participação do utilizador no processo de criação da solução e valoriza a sua presença nas diversas etapas de validação.

3. ÁREAS DE APLICABILIDADE

A metodologia de design thinking começou a ser conhecida e utilizada num contexto de desenvolvimento de produtos, mais especificamente, num contexto de desenvolvimento de aplicações de *software*. No entanto, dada a sua abrangência, a aplicação da mesma tem sido sugerida e comprovada noutros contextos. Esta metodologia, mais do que uma forma de chegar a soluções ótimas, pretende apresentar uma forma sistemática de considerar um problema e, a partir daí, encontrar soluções mais adequadas. Assim, podem considerar-se como base os problemas mais diversos, como o desenho de um novo produto, a definição de um processo de negócio, a otimização de um serviço ou mesmo a definição de uma política.

Só segundo esta perspetiva se podem compreender os exemplos de utilização de design thinking em áreas tão diversas como a definição de políticas governamentais, a definição de regras de acesso a cuidados de saúde ou mesmo a definição de sistemas de distribuição de água.

A utilização de design thinking em áreas tão diversas é possível sempre que o problema em causa tenha uma forte dependência humana, podendo ser de alguma forma caracterizado como "*human centric*". Se em termos de desenvolvimento de produtos falamos em utilizadores, na definição de políticas falamos em cidadãos, constituintes. Ambas as áreas reconhecem as pessoas como a unidade básica de decisão que recebe influências de outros fatores, mas que no final é quem delibera. Como a metodologia centra a sua análise no estabelecer de empatia com os utilizadores, cidadãos ou qualquer outra designação, a sua aplicação torna-se possível e real.

Outro aspeto importante que permite esta vasta utilização é a dependência do contexto. Tanto a aceitação de novos produtos, como a aceitação de novos processos ou de novas políticas vai depender muito do contexto em que a pessoa, como unidade básica de decisão, se insere. Para se entender este contexto é necessário utilizar uma abordagem multidisciplinar, conjugando áreas muito diferentes como a economia, psicologia, engenharia, antropologia ou mesmo a biologia. A metodologia de design thinking vem endereçar exatamente esta necessidade de foco no contexto e integração destes conhecimentos dispersos, uma vez que introduz métodos de recolha, análise e entendimento de relacionamentos, e até de identificação de resultados não expectáveis.

Finalmente, o último aspeto que promove uma utilização mais vasta é a natureza iterativa e adaptativa que o design thinking apresenta. É virtualmente impossível acertar à primeira numa solução ótima para problemas complexos, pelo que têm de ser colocados em prática métodos que permitam a experimentação, iteração e evolução da solução que permitam entender o problema, mas também adequar essa mesma solução às várias necessidades intercruzadas que possam existir.

Tal como produtos falhados, assiste-se igualmente à definição de políticas com pouco sucesso, sendo que grande parte delas não consideraram as verdadeiras necessidades dos seus "utilizadores", nem tão pouco foram sujeitas a processos de validação e de recolha de feedback.

A utilização de design thinking está a ser alargada a outros domínios que não apenas o desenvolvimento de produtos, sendo que apresenta já alguns casos de sucesso, os quais devem ser considerados como um incentivo a que esta utilização seja continuada e até reforçada.

4. CASO DE APLICAÇÃO PRÁTICA

Na PT Inovação decorreu, no ano de 2012, através do evento denominado Tarde de Inovação, um exemplo de aplicação transversal do design thinking. Este pretendeu ser um momento de reflexão em que se conjugou a aprendizagem de algumas técnicas desta metodologia, a familiarização com a dinâmica e as etapas de um processo criativo, a prática do processo de criação coletiva, a assimilação das linhas orientadoras do Rotas (Roteiro Tecnológico) e o aumento da motivação para a produção de novidade.

As técnicas utilizadas incidiram sobre produtos PT Inovação, selecionados de acordo com as linhas de inovação do Rotas (IAM, ONT, Parknow, Pocket Account, Secret Story, Telemetria, Medigraf).

O design thinking preconiza a introdução de métodos de design e a sua cultura em campos mais afastados do design tradicional, como a engenharia de produto e processo, o desenho de regras de gestão de negócios, etc. Para tal, procura considerar em

>>>

simultâneo necessidades humanas e novas perspetivas de vida com qualidade, material disponível e recursos tecnológicos, constrangimentos e oportunidades para um projeto ou negócio.

Um design thinker caracteriza-se pela sua capacidade de descobrir novas oportunidades, de pensar em variedade, possibilidades e múltiplas perspetivas, de considerar situações que se distanciem de estereótipos e preconceitos.

Em grupos multidisciplinares e utilizando (i) brainstorming e entrevistas ao gestor de produto (para criação de empatia), (ii) mapas de empatia para criação de um ponto de vista do utilizador, questionar os pressupostos do produto e receber novas perceções e conhecimento (ainda na fase de empatia) e (iii) mapas mentais para visualização do entendimento do grupo acerca do produto em causa (na fase de ideação), os colaboradores da PT Inovação estiveram em contacto com técnicas de design thinking para a componente de divergência, na qual a preocupação é a de criação e disponibilização de escolhas (liberdade para geração de ideias).

Como resultados, para além da difusão da metodologia internamente, dando a conhecer algumas técnicas que podem ser introduzidas no dia-a-dia de trabalho dos diferentes projetos que decorrem na empresa, resultou ainda a inclusão em roadmap de 2% das ideias que surgiram na Tarde de Inovação.

5. CONCLUSÕES

O design thinking é mais uma ferramenta que coloca as técnicas de design e a criatividade ao serviço da inovação, ultrapassando portanto a componente estética [3]. Dos procedimentos que compõem esta metodologia, podem enumerar-se os que mais se destacam no caminho para a inovação: observar atentamente e criar empatia, pensar virtualmente e integrar todos os sentidos, identificar padrões e reinventá-los, negar estereótipos, estabelecer e recombinar novas ligações, pensar em analogias e trabalhar em equipa. [2]

Vemos com frequência produtos que não são os primeiros a chegar ao mercado, mas que são aqueles que verdadeiramente nos atraem emocionalmente e funcionalmente. [3]

A inovação parte da nossa capacidade de observarmos as pessoas e analisarmos a sua perspetiva em detrimento da nossa. Para que a inovação ocorra, para que as ideias ganhem vida, é necessário mudar drasticamente a forma de pensar. É uma disciplina completamente nova. O design thinking é uma proposta nesta direção, utilizando a sensibilidade e métodos do design para responder às necessidades das pessoas, considerando o que é tecnologicamente e estrategicamente viável. O objetivo final é criar valor para o cliente e novas oportunidades de mercado. "A abundância tem satisfeito, e até mesmo sobre satisfeito, as necessidades materiais de milhões - aumentando a importância da beleza e da emoção, e acelerando a busca dos indivíduos por aquilo que faz sentido." Estas experiências não se irão traduzir em produtos simples, correspondendo a combinações complexas de produtos, serviços, espaços e informação. [3]

As técnicas de design thinking não são transversais às equipas de trabalho na PT Inovação. Os resultados da sua divulgação resultam ainda num impacto reduzido para a dimensão da empresa. A sua aplicação em projetos pontuais tem uma expressão diferente, com importância assumida pelas equipas técnicas. Acredita-se todavia que a utilização das técnicas aqui descritas fomentará uma aproximação ao utilizador dos produtos e serviços implementados, uma maior abertura e atenção à envolvente, com criação de valor num mercado globalizado, em que a análise de valor acrescentado é cada vez mais guiada por fatores emocionais.

A inovação cai muitas vezes entre uma grande ideia e a capacidade de a colocar em prática. O design thinking é uma metodologia que disponibiliza técnicas que auxiliam nesta transição, acrescendo da proximidade que é trabalhada com o cliente/utilizador final.

REFERÊNCIAS

[1] Hasso Platner Institute of Design at Stanford
[2] Processos Criativos e Design Thinking nas Empresas, Katja Tschimmel
[3] Brown T., Design Thinking, Harvard Business Review, June 2008
[4] Algoso D., A view from the cave, http://www.aviewfromthecave.com/2013/06/politics-and-design-thinking-more-in.html
[5] Brown T., Wyatt J., Design Thinking for Social Innovation, http://www.ssireview.org/articles/entry/design_thinking_for_social_innovation/
[6] Preece J. (2000), *Online Communities: Supporting Sociability, Designing Usability*, Wiley

CVS DOS AUTORES

Inês Oliveira, obteve o Mestrado em Redes de Informação pela Carnegie Mellon University e Universidade de Aveiro em 2008 e a Licenciatura em Engenharia de Computadores e Telemática pela Universidade de Aveiro em 2004. Ingressou na Portugal Telecom em 2004, fazendo parte da equipa de sistemas e infraestruturas de redes, tecnologias IP. Em 2009 ingressou as equipas de protocolos e aplicações para gestão de elementos de rede e de revisão dos processos de gestão de programas e projetos. À vertente de consultoria de redes seguiu-se a integração na equipa de Gestão de Inovação, em 2011, responsabilizando-se pela coordenação da Inovação na PT Inovação e gestão do Sistema de Investigação, Desenvolvimento e Inovação (IDI). Faz parte da equipa do projeto transversal de Contextual Design na PT Inovação.

Lúcia Moreira, Doutoranda em Informação e Comunicação em Plataformas Digitais na Universidade de Aveiro e na Universidade do Porto e licenciada em Novas tecnologias da Comunicação pela Universidade de Aveiro. Trabalha no Departamento de Comunicação e Imagem da PT Inovação, onde é responsável pela área de Usabilidade e User Experience. Formadora do curso Formação Pedagógica para eFormadores. Autora e coautora de artigos e comunicações na área da comunicação multimédia, usabilidade, arquitetura de informação, eLearning e novos contextos de aprendizagem. Gestora de projetos de investigação nas áreas de Comunicação e eLearning. Faz parte da equipa do projeto transversal de Contextual Design na PT Inovação.

Nuno Alexandre Seixas licenciou-se em Engenharia Informática pela Universidade de Coimbra em 2004. Em 2008 obteve o grau de Mestre em Engenharia de Software pela Carnegie Mellon University e Universidade de Coimbra. Neste momento é Gestor de Projetos na PT Inovação, sendo responsável pelos projetos de melhoria interna. É Instrutor certificado de CMMI para Desenvolvimento, v1.3. Faz parte da equipa do projeto de Contextual Design na PT Inovação. Fez parte da equipa de desenvolvimento de software para a área de Informática Médica, nomeadamente da plataforma de Telemedicina – Medigraf. Fez parte da equipa do projetos de investigação – MyHeart e mCare, projetos europeus na área da Informática Médica no contexto do Centro de Informática e Sistemas da Universidade de Coimbra.

02 Aplicações de Visão por Computador


FAUSTO DE CARVALHO JOSÉ ALBERGARIA

PALAVRAS CHAVE
Computer Vision, Realidade Aumentada, Mobilidade, Aplicações, Meo, Move4Health, Online-Gym, MarvIN

A visão por computador é um tema vasto e objeto de múltiplas linhas de experimentação e aplicação, intercetando áreas tão díspares como robótica, medicina, controlo industrial e segurança.

Este artigo caracteriza sumariamente o domínio, para enquadrar uma geração emergente de aplicações pessoais, mobile e desktop, que recorrem a análise da imagem recolhida com câmaras e periféricos afins de baixo custo para a disponibilização de novas funcionalidades de interação baseada no processamento digital de imagem, cor e movimento.

São abordados os principais conceitos e técnicas, com ênfase para alguns trabalhos de inovação exploratória atualmente em curso e perspetivando a sua aplicação em cenários de produção.

1. INTRODUÇÃO

O mundo das tecnologias de informação e comunicação continua a registar evolução a um ritmo alucinante, potenciada por conteúdos multimédia de elevada qualidade, acessíveis em banda larga cada vez mais generalizada, através de terminais móveis poderosos e versáteis, imersos numa mescla de redes de sensores. É o advento da banalização dos serviços de Realidade Aumentada em tempo real, potenciando ambiciosos e interessantes cenários em áreas tão díspares como turismo, saúde ou *smart cities*.

Indo mais e mais longe, aplicações que "ouvem", "vêm" e "sentem" cada vez mais irão ampliar o poder e o âmbito dos processos cognitivos e perceção sensorial humana, numa abordagem que é já referida como Cognição Aumentada [1].

A investigação e desenvolvimento prosseguem atualmente em torno de tecnologias com elevado grau de maturidade, tais como processamento digital de imagem, *scanning*, visualização e impressão 3D, mas também trilhando novos caminhos nos domínios das nanotecnologias, neurociências e engenharia biomédica, enquadrando e desafiando o nosso dia-a-dia por exemplo com biossensores, implantes de retina, BCI (*brain-computer interfaces*) e dispositivos microscópicos que, apesar de cada vez mais credíveis, cruzam claramente as fronteiras da ficção e dão asas à nossa capacidade de imaginar um futuro ainda mais conectado, com oportunidades e desafios para a introdução de novos serviços.

A computação baseada no processamento de imagem é uma componente chave para muitas dessas aplicações emergentes: a nossa compreensão do contexto e ambiente depende fortemente da visão e isso também acontece frequentemente nos sistemas de cognição aumentada. É assim natural que se tenha vindo a registar um forte aumento, maturidade e interesse no recurso a técnicas de visão por computador (*computer vision*) para resolver os mais diversos problemas de engenharia.

2. A EVOLUÇÃO DA VISÃO POR COMPUTADOR

A visão por computador pode ser definida como a área de conhecimento que endereça aquisição, processamento, análise e interpretação de imagens do mundo real com o objetivo da tomada de decisões autónomas. O *know-how* necessário ao domínio e aplicação das técnicas de visão por computador é alicerçado num forte conhecimento teórico dos modelos e ferramentas matemáticas que estruturam todos os princípios aplicáveis, mas necessita igualmente de todo um conjunto de valências específicas de informática e ciências da computação ao nível dos algoritmos, plataformas, *software* e *hardware* que permitam a concretização das funcionalidades com a eficácia requerida pelos múltiplos cenários de aplicação (figura 1).

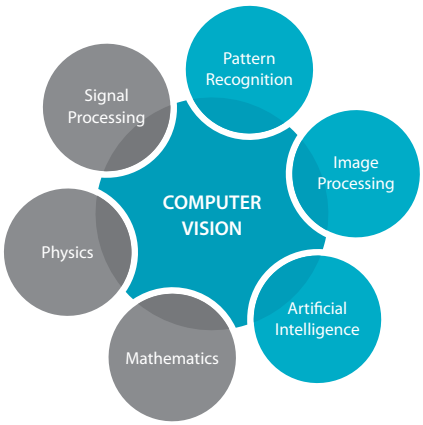


Figura 1. Âmbito da Visão por Computador (fonte: maxembedded.com)

>>>

Alguns dos primeiros trabalhos relevantes em visão por computador surgiram do lado da robótica, explorando a possibilidade de criação de um robô capaz de adquirir *input* visual do mundo real e decidir com base na interpretação das imagens percebidas. Inicialmente acreditava-se que a interpretação seria uma tarefa acessível, mais ao alcance e realizável que os pesadíssimos algoritmos de tomada de decisão e aprendizagem que a inteligência artificial já vinha a explorar. Todavia a reprodução ou recriação da complexidade e capacidade da visão humana veio a revelar-se uma tarefa hercúlea, à medida que se foi sabendo mais e foram sendo identificados novos aspetos e dificuldades. Apesar de tudo, o caminho que ainda hoje se continua a trilhar conduziu já à criação de ferramentas que permitem tarefas impressionantes aplicadas nas mais variadas áreas da ciência, industria e engenharia.

Hoje em dia recorre-se com sucesso a visão por computador aplicada por exemplo em robôs industriais para controlo de processos, na navegação em veículos autónomos, em videovigilância, em interação homem-máquina e em imagiologia médica, em subdomínios que incluem deteção de eventos, *video tracking*, reconhecimento de padrões, *machine learning* e estimação de movimento, entre outros.

Alguns dos conceitos teóricos basilares em que assenta a visão por computador são-nos familiares desde há muito.

É o caso da perspetiva e da representação do mundo real numa superfície 2D, conceito introduzido no século XV por desenhos de Filippo Brunelleschi ("*Construzione Legittima*") e posteriormente desenvolvido por Piero della Francesca ("*De Prospectiva Pingendi*") e Leonardo da Vinci. A par do uso da perspetiva como ferramenta artística, foi sendo desenvolvido um intenso estudo matemático na área da geometria (p.ex. Desargues no século XVII), que dotou artistas e cientistas de ferramentas que permitiram uma representação exata das distâncias, posições e proporções e estas são ferramentas extensivamente usadas atualmente em algoritmos para ajuste de perspetiva, cálculo de distâncias ou posições relativas em imagens 2D, cruciais em algumas das tarefas associadas a visão por computador.

Outro marco histórico importante é o modelo de reflexão difusa de luz que resulta da Lei de Lambert (1760) (também conhecida por Lei de Beer ou Lei de Beer-Lambert-Bouguer, por ter sido descoberta empiricamente e independentemente pelos três matemáticos no século XVIII) e que continua a ser bastante usado, nomeadamente em reconhecimento de formas e deteção de movimento.

A importância da matemática na visão por computador é de facto inequívoca e a compreensão do conhecimento relevante gerado nos últimos três séculos ao nível da álgebra, estatística, geometria e cálculo vetorial tem um valor inquestionável para a correta definição de cada problema, estabelecimento dos algoritmos adequados e sua concretização aplicacional.

Larry Roberts, um dos criadores da ARPANET (a par de Leonard Kleinrock, Vincent Cerf e Robert Kahn) é igualmente considerado um dos percursores desta área e autor de um dos primeiros artigos sobre visão por computador. No âmbito da sua tese de PhD no MIT, em 1963 publicou "*Machine Perception of Three-Dimensional Solids*" [2], onde foi apresentado um método para, a partir de uma imagem de sólidos geométricos, detetar quais os sólidos que compõem a imagem e calcular a sua representação a partir de um ponto de vista alternativo. A sua abordagem tornou-se clássica e os seus algoritmos ainda hoje são bastante utilizados, apesar de curiosamente serem escassas as referências nos conteúdos bibliográficos sobre o autor, certamente devido à extrema importância dos seus contributos para a área de *networking* e da Internet.

O trabalho de Roberts foi a tónica para as décadas que se seguiram, onde múltiplos estudos foram feitos na tentativa de deteção das estruturas 3D presentes nas imagens, inicialmente através da deteção de linhas e curvas e da sua análise topológica, mais tarde através do uso de sólidos de revolução e *cilindros genéricos* - técnicas ainda amplamente usadas para reconhecimento de objetos. Ainda numa tentativa de explorar as características presentes nas imagens para uma melhor interpretação do seu conteúdo, foi explorada uma abordagem qualitativa das intensidades e variações das sombras e a sua relação com a forma dos objetos. Outra inovação no reconhecimento de objetos, foi a utilização de contornos para a inferência da forma 3D.

Os anos 80 evidenciaram uma viragem para a otimização e sofisticação das ferramentas matemáticas aplicadas a visão por computador, com grande enfoque na deteção de orlas e formas. Foi conseguida uma unificação das *frameworks* matemáticas, através de uma formulação mais cuidada dos problemas.

Já nos anos 90 registou-se progresso assinalável, por exemplo com a introdução da técnica "*structure from motion*" através da qual objetos 3D são

reconstituídos a partir de sequências de imagens 2D do objeto em movimento. Começou a ser usada "*physics-based vision*", uma nova abordagem em que os objetos são modelados mediante observação das suas propriedades eletromagnéticas, nomeadamente radiância e comprimento de onda do espectro visível. Foi ainda nesta década que se consolidou a interação entre visão por computador e computação gráfica com a manipulação de imagens reais pela interpretação do seu conteúdo.

A deteção baseada em *features* foi introduzida na década de 2000. Esta técnica permite extrair características visuais relevantes de objetos ou cenas e usá-las na sua deteção em imagens reais. Também a área de reconhecimento facial teve grande evolução na década passada, com a criação de algoritmos que permitem a deteção e identificação de rostos com baixa taxa de falsos positivos, aplicados a sistemas de segurança de grande desempenho.

3. VISÃO POR COMPUTADOR: ALGUNS CONCEITOS BÁSICOS

Há todo um conjunto de tarefas mais ou menos complexas que é necessário levar a cabo para se conseguir "ver", na lógica da visão computacional. Ao longo dos próximos parágrafos propomo-nos abordar sumariamente alguma dessas operações e terminologia associada, que não são mais do que um mero ponto de partida para um domínio extremamente vasto e exigente.

Uma das tarefas cruciais é a captura e armazenamento de imagem. Uma imagem é obtida a partir dos drivers dos dispositivos de vídeo e armazenados em estruturas de dados, tipicamente matrizes bidimensionais, que armazenam os valores relativos a cada um dos pixéis que compõem a imagem. As imagens têm uma componente de altura e largura que são mapeados nas linhas e colunas da matriz. Em cada ponto da matriz, é armazenado o valor de cor do pixel. Se a imagem é a preto e branco (binário), será armazenado "1" ou "0" em cada ponto. Se a cor é dada em escala de cinzentos, vai ser armazenado um valor no intervalo da escala de cinzentos escolhida. Se for uma imagem a cores, a matriz será um *array* multidimensional que irá armazenar um conjunto de valores que depende da representação de cores utilizada (e.g. RGB - *red green blue*). Assim, para cada pixel da imagem vai ser incluída informação de cada componente de cor, bem como eventualmente um valor adicional de alfa (transparência), sendo típico chegar-se a uma representação que necessita de 24 bits para cada pixel, ou 32 bits se incluir alfa. É fácil concluirmos que para imagens a cores com resolução elevada, bastante comuns nos dias de hoje, as necessidades de processamento e memória podem ser gigantescas devido ao enorme volume de dados a armazenar e processar.

A partir destas estruturas de dados, é possível realizar um conjunto de ações preliminares que visam preparar a imagem para transformação e análise. Essas ações endereçam operações simples, tais como redimensionamento para reduzir a quantidade de dados associados à imagem ou conversão para um formato específico, mas podem igualmente envolver alguma semântica, por exemplo a seleção de uma ROI (*Region Of Interest*) para processamento, sem necessidade de lidar com a totalidade da imagem e consequentemente aligeirando e otimizando o processamento subsequente.

As técnicas a utilizar incluem deteção de cor e deteção de orlas (*edge detection*). Através da deteção de cores, é possível selecionar um conjunto de pixéis que correspondem a uma determinada gama de cores, o que pode ser usado por exemplo para recuperar um objeto ou um contorno. *Edge detection* permite a identificação de descontinuidades, que podem ser interpretadas como linhas de fronteira entre diferentes realidades, podendo ser mapeadas em formas, objetos, rostos ou regiões. Uma das mais-valias desta técnica é a faculdade de fazer emergir contornos presentes na imagem, ignorando ruído visual e/ou informação potencialmente ilusória.

Como extensão à deteção de orlas, podemos considerar a deteção de formas, o que implica a deteção de uma linha fechada, delimitando uma cena ou um objeto. Através da aplicação de operadores matemáticos a orlas, é possível detetar formas, verificando a presença de linhas retas, curvas e ângulos. Uma aplicação típica para este tipo de processamento é o reconhecimento de cartas: uma vez efetuado o reconhecimento do contorno retangular, um algoritmo de deteção de formas pode ser usado de seguida para contar o número de objetos (por exemplo "copas") presentes e assim identificar qual a carta (por exemplo "5 de copas").

Na maioria dos casos reais em que queremos recorrer a aplicações de visão por computador, o objeto do reconhecimento é uma composição complexa, por exemplo uma casa ou um jardim, ou até mesmo uma cena composta por vários objetos que interagem entre si. A técnica mais simples para trabalhar uma cena complexa em imagens do mundo real é *Template Matching*. Nesta técnica, o algoritmo mais básico e direto é a busca por correspondência pixel

>>>

a pixel entre a imagem real e um *patch*. Esta pesquisa é conseguida fazendo deslizar o *patch* através da imagem, um pixel de cada vez, avaliando a correlação entre o *patch* e os pixéis da imagem (convolução). Diferentes métodos matemáticos podem ser utilizados para calcular e avaliar a correlação. A decisão sobre se o *patch* está presente ou não irá depender da função usada no cálculo da correlação e dos parâmetros de decisão. Se o máximo da função de correlação for maior que o parâmetro (limiar) fornecido, assume-se a presença do *patch* e a sua posição é a posição desse máximo da imagem.

Este método é muito vulnerável a variações de iluminação, ruido visual ou perspetiva. Também é computacionalmente exigente, o que o torna inadequado para a deteção de vídeo em tempo real, nomeadamente quando usado em dispositivos móveis. A carga computacional é elevada devido ao processo de convolução que, apesar de simples, é pesado devido ao elevado volume de dados a processar.

É possível otimizar o processo identificando ROI na imagem que permitam extrair características distintivas, que possam ser usadas para detetar inequivocamente o *patch* desejado. Apesar do aumento de complexidade, normalmente obtém-se ganho significativo de desempenho.

Outro problema comum quando se trabalha com vídeo em tempo real é a ocorrência de oclusão de um número variável de pixéis, o que dificulta a deteção. Uma forma de superar esta situação é a utilização da técnica de *Feature Detection and Matching*. Uma *feature* pode ser um angulo, orla ou *blob* - *blob* são regiões que emergem de uma imagem como tendo uma característica distintiva como cor ou brilho. Os *blobs* podem ser reconhecidos através da aplicação de operadores matemáticos e/ou transformações da imagens que enfatiza propriedades estatísticas locais. Existem muitos algoritmos diferentes para extração e descrição de *features* que tipicamente estão segmentados em três etapas diferentes: extração, descrição e comparação. A extração de características implica a análise de um *patch* e é a determinação de quantas *features* estão presentes e quais relevam melhor as suas características distintivas. Após a extração, as *features* são descritas quanto à sua natureza (angulo, *blob*, …), a sua posição dentro da imagem e a relação entre elas, o que pode ser visto como a criação da assinatura visual do *patch*. O último passo é a comparação entre a assinatura do *patch* e a assinatura da imagem fonte, por forma a calcular a correlação entre elas. Se a correlação for superior a um limiar predefinido, assume-se a existência do *patch* na imagem.

Na discussão sobre qual o melhor algoritmo de deteção, parâmetros como o tempo médio de deteção, número de *features*, qualidade das *features* e qualidade do rastreamento são fundamentais para decidir qual se adapta melhor a cada problema. A importância de cada um desses parâmetros depende da natureza e das circunstâncias da aplicação do algoritmo. Em alguns contextos, pode valorizar-se a redução de falsos positivos, apostando-se em algoritmos complexos e pesados, mas com menos falhas. Noutras situações, como no caso dos dispositivos móveis, podemos querer trocar precisão por economia de recursos.

4. TECNOLOGIA PARA VISÃO POR COMPUTADOR

À medida que a visão por computador foi ganhando relevância e sendo reconhecida como uma tecnologia-chave, foram ficando disponíveis bibliotecas e *frameworks*, seguindo as principais tendências do ecossistema de desenvolvimento de aplicações. Criados e mantidos por empresas de renome tais como Intel ou Qualcomm, por *startups* ou até mesmo por programadores individuais, de uma forma geral todos esses ambientes permitem a integração de ferramentas e funcionalidades de visão computacional em aplicações, seja para uso em desktop ou móvel, para ambientes industriais ou mesmo de uso doméstico.

A biblioteca de visão computacional inquestionavelmente mais popular é OpenCV (Open *Source Computer Vision*). A iniciativa OpenCV foi oficialmente lançada em 1999 pela Intel Research, enquadrada num conjunto de projetos que endereçavam aplicações de *ray tracing* e 3D *display walls*. O seu objetivo era assumidamente promover o avanço da investigação em visão por computador, mas também disponibilizar uma infraestrutura de código otimizado reutilizável, de forma a disseminar o conhecimento da área no seio da comunidade de desenvolvedores e investigadores e contribuir para a introdução de aplicações comerciais de visão por computador baseadas numa biblioteca poderosa disponível gratuitamente. O sucesso do OpenCV é evidente: desde 2012 entregue à fundação sem fins lucrativos opencv.org, é suportado pelas *startups* Willow Garage e Itseez, já registou mais de 6 milhões de *downloads* e estima-se que é usado por cerca de 50 mil desenvolvedores. Escrito em C/C++ otimizado, pode tirar partido de processamento *multicore*, possui interfaces em C++, C, Python e Java, suportando Windows, Linux, Mac OS, iOS e Android e está disponível sob licença BSD (gratuito tanto para uso académico como comercial). [3]

A primeira experiência da PT Inovação com visão por computador foi apresentada no 10º Encontro Português de Computação Gráfica em 2001 [4], consistindo numa avaliação da utilização de *webcams* de baixo custo para interação homem-máquina. O desenvolvimento do protótipo foi baseado na biblioteca OpenCV e Direct Show em ambiente Windows (figura 2).

Figura 2. Protótipo de IHM por deteção de movimento (2001)

Anos mais tarde, voltámos a recorrer a OpenCV como base para o desenvolvimento de uma aplicação IPTV para reconhecimento facial, tendo o protótipo meoID sido demonstrado com elevado sucesso durante o evento Sapo Codebits 2008 (figura 3).

Figura 3. Protótipo meoID (Codebits 2008)

A Qualcomm é outro dos *players* que tem vindo a disponibilizar plataformas de visão por computador, nomeadamente para aplicações mobile. O seu SDK FastCV, disponível para Windows, OS X e Linux, dá acesso a uma biblioteca com funcionalidades de reconhecimento de gestos, deteção, *tracking* e reconhecimento facial, reconhecimento e *tracking* de texto e realidade aumentada, que podem correr em qualquer processador ARM e especialmente otimizadas para processador Snapdragon.

Mais recentemente, a Qualcomm introduziu a plataforma Vuforia, destinada em particular a aplicações de realidade aumentada. Disponível para Android, iOS e Unity, permite a deteção de imagens e reprodução de conteúdo digital em vídeos de tempo real, em áreas tais como entretenimento, educação, jogos e publicidade. [5]

BoofCV é outro bom exemplo do dinamismo da indústria da visão por computador. Desenvolvida por Peter Abeles, é uma biblioteca Java *open source* para visão computacional em tempo real. Ainda no início do desenvolvimento, é uma tentativa individual de portar ferramentas de visão por computador para ambientes populares de produção, através da integração da biblioteca com IDEs amplamente usados (Eclipse) e incorporação de bibliotecas matemáticas. [6]

É ainda de referir SimpleCV, uma *framework open source* Python (com um *port* em curso para Javascript) para desenvolvimento rápido de aplicações de visão por computador, que integra o acesso a várias bibliotecas poderosas, p.ex. OpenCV, mas abstraindo a complexidade dos conceitos de base, sendo por isso referida como "*computer vision made easy*". [7]

5. VISÃO POR COMPUTADOR EM CONTEXTO DE MOBILIDADE

Tal como foi sendo discutido ao longo deste artigo, são múltiplas as áreas de aplicação de visão por computador com relevância para a atividade da PT Inovação e empresas PT, não sendo por isso de estranhar a ocorrência de diversas iniciativas de experimentação e inovação exploratória neste domínio.

É o caso do projeto Move4Health, desenvolvido por uma equipa multidisciplinar do Instituto de Telecomunicações do Porto (Porto Interactive Center e Instituto Politécnico do Porto) no contexto do Plano de Inovação 2013-2014, em parceria e com financiamento da PT Inovação. Tem como objetivo realizar experimentação com tecnologias de captura de movimento em tempo real sem recurso a marcadores (*markless mocap*) e ambientes virtuais 3D interativos no contexto da avaliação clínica e potencial de reabilitação de capacidade motora (*large motor movements*) e validação através da criação de provas de conceito baseadas em abordagem "*serious game*", incluindo uma avaliação comparativa entre a utilização de um sistema profissional (Organic Motion) [8] e um periférico de consumo (Microsoft Kinect) [9].

Outro projeto relevante é o Online-Gym, realizado pelo polo do INESC TEC na Universidade de Trás-os-

>>>

-Montes Alto Douro (UTAD), igualmente no contexto do Plano de Inovação 2013-2014, em parceria e com financiamento da PT Inovação. O objetivo é a experimentação com tecnologia de captura de movimento em tempo real (*mocap*) de baixo custo (Microsoft Kinect), para sincronização e monitorização de atividade física do tipo ginástica em grupo, efetuada *online* por utilizadores isolados, geograficamente dispersos ou com mobilidade reduzida, incluindo a criação de protótipo de serviço imersivo 3D em ambiente OpenSimulator / Second Life [10].

Todavia foi na área das ferramentas de realidade aumentada em contexto de mobilidade para suporte à operação, manutenção e gestão que foram identificadas oportunidades mais significativas e, por isso, centrados os esforços recentes de introdução de aplicações baseadas em visão por computador.

Foi esse o cenário-alvo selecionado para a tese do Mestrado em Comunicação Multimédia da Universidade de Aveiro (DeCA) com o tema "Visualização Interativa em contexto de *Computer Vision*". Foram exploradas questões de interatividade associadas à visualização de informação tridimensional enquanto elemento de realidade aumentada em aplicações mobile baseadas em visão por computador, tendo sido desenvolvido o protótipo funcional MarvIN, que permitiu avaliar, validar e demonstrar algumas das conclusões do trabalho de investigação (figura 4) [11].

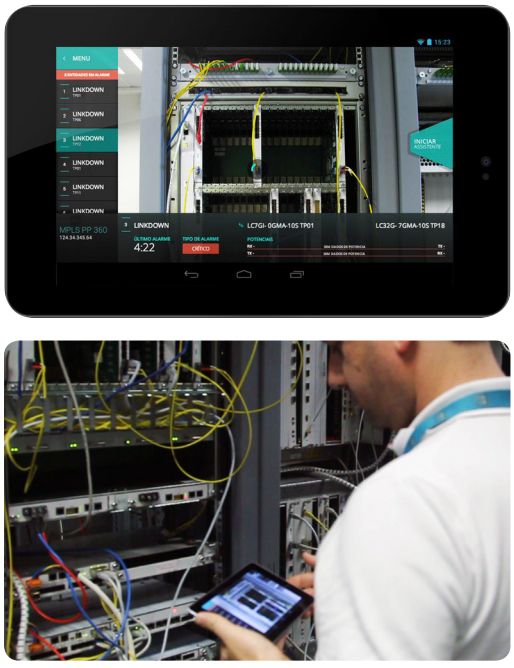


Figura 4. MarvIN - Realidade Aumentada em Contexto de Mobilidade (2013)

Também no âmbito do Plano de Inovação 2013-2014, merece destaque o projeto "Realidade Aumentada em Contexto de Mobilidade - Computer Vision" conduzido pelo Instituto de Sistemas e Robótica da Universidade de Coimbra, em parceria e com financiamento da PT Inovação. Trata-se de avaliação do estado da arte e experimentação para a realização de protótipo de uma ferramenta de Realidade Aumentada através de reconhecimento de padrões por análise da imagem da câmara de um dispositivo móvel (*smartphone* ou *tablet*).

O objetivo é a disponibilização de funcionalidades que possam ser utilizadas em tarefas de operação, manutenção e gestão de salas técnicas e bastidores, incorporadas em ferramentas associadas aos sistemas de gestão e cadastro, para identificação visual dos equipamentos e da sua localização no bastidor, correspondência com cadastro, *overlay* de informação multimédia contextualizada (e.g. numeração de leds, alarmística) e identificação do estado de leds (ligado/desligado) e conetores (livre/em uso) (figura 5).

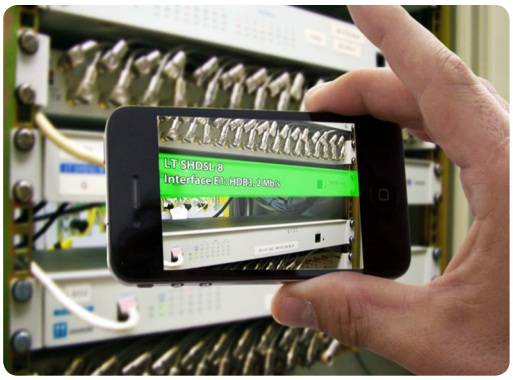


Figura 5. Realidade Aumentada em Contexto de Mobilidade (mockup) (2013)

Os desafios que se colocam à aplicação de tecnologia de visão por computador a este tipo de cenários são múltiplos, em face da complexidade intrínseca do contexto, a que acrescem questões físicas de espaço, iluminação e ruído visual, entre outras, mas a oportunidade de gerar ganhos de produtividade e acrescentar valor percetível para os clientes faz-nos ter a ambição de conseguir dominar este tópico e caminhar com determinação no sentido de, em breve, começarmos a incorporar visão por computador no ensaio e teste de equipamentos, no suporte às operações, na formação e até mesmo nos conteúdos e manuais multimédia fornecidos aos nossos clientes.

REFERÊNCIAS

[1] Rahul Swaminatham, Fausto de Carvalho, "Augmented human cognition and the impact on the network", Eurescom mess@ge 1/2012

[2] Larry Roberts, "Machine Perception of Three-Dimensional Solids", MIT 1963, http://www.packet.cc/files/mach-per-3D-solids.html

[3] OpenCV - Open Source Computer Vision Library, http://opencv.org/

[4] Ricardo Ferreira, Bernardo Cardoso, Fausto de Carvalho, "Interacção Homem-Máquina por detecção de movimento", 10º Encontro Português de Computação Gráfica, Lisboa 2001

[5] Qualcomm Developer Network, http://developer.qualcomm.com/mobile-development/mobile-technologies/

[6] BoofCV, http://boofcv.org/

[7] SimpleCV, http://simplecv.org/

[8] Organic Motion OpenStage, http://www.organicmotion.com/applications/animation/

[9] Microsoft Kinect, http://www.microsoft.com/en-us/kinectforwindows/

[10] OpenSimulator, http://opensimulator.org/wiki/Main_Page

[11] Ricardo Nascimento, "Visualização Interativa em contexto de Computer Vision", Mestrado em Comunicações Multimédia, Universidade de Aveiro, 2013

CVS DOS AUTORES

Fausto de Carvalho, é licenciado em Engenharia Eletrónica e Telecomunicações pela Universidade de Aveiro. O seu percurso no CET e PT Inovação está fortemente ligado à área da Interatividade e Tecnologia Multimédia, tendo exercido funções de gestão e participado em múltiplos projetos de I&D nacionais e internacionais, com especial envolvimento na introdução do serviço IPTV Meo. É co-autor de várias patentes e de numerosas comunicações e artigos técnicos. Enquanto consultor tecnológico sénior, atualmente centra a sua atividade na exploração e demonstração de novas tecnologias, conteúdos e aplicações emergentes, em múltiplos contextos de convergência, conetividade e mobilidade, nomeadamente coordenando e participando em projetos de inovação exploratória em parceria com instituições universitárias.

José Albergaria, é licenciado em Engenharia Eletrónica e Telecomunicações pela Universidade de Aveiro. Começou o seu percurso na PT Inovação ligado aos projetos europeus onde trabalhou em áreas relacionadas com *User Generated Services* (UGS) e plataformas de telecomunicações. Posteriormente migrou para área mobile onde adquiriu competências no sistema operativo Android tendo participado em múltiplos projetos de I&D, com especial envolvimento na fusão de aplicações Android com outras tecnologias nomeadamente: IPTV, *Near Field Comunications* (NFC), Eletroencefalografia (EEG), *Computer Vision* (CV) e indoor location. Esteve presente também no desenvolvimento de várias aplicações mobile associados a produtos e serviços PT.

03 Bases de Dados NoSQL: Taxonomia e Tradeoffs de Desempenho, Disponibilidade e Consistência de Dados



MANUEL GONÇALVES MÁRIO MOREIRA MIGUEL BISCAIA

PALAVRAS CHAVE
Consistência atómica, consistência eventual, latência, *throughput*, alta-disponibiidade, fiabilidade

O desenho e implementação de sistemas contemporâneos de classe empresarial, advoga a crescente necessidade do reequacionamento dos equilíbrios de disponibilidade, consistência e desempenho, numa perspetiva de que a ideologia "*one size fits all*" associada às bases de dados relacionais é em diversos casos onerosa e indutora do subaproveitamento de recursos computacionais. Neste sentido, emergiu um conjunto vasto de paradigmas e soluções de gestão de dados, comummente designados por NoSQL (*Not Only SQL*). Neste artigo, são introduzidas as caraterísticas funcionais e arquiteturais dos principais paradigmas, seguindo uma abordagem de apresentação de casos de uso, e de indicadores qualitativos e quantitativos de exemplos reais de bases de dados NoSQL.

1. MOTIVAÇÃO

A questão primordial que se coloca aquando da discução do tema NoSQL é qual a motivação do surgimento destas soluções de gestão de dados? Obviamente que há uma multitude de argumentos, dos quais se salientam os seguintes vetores:

- **Volume** - de facto têm-se assistido a um aumento exponencial da quantidade de informação a gerir, especialmente num contexto de um operador de comunicações. As RDBMS tradicionais foram idealizadas para cenários OLTP com volume de informação controlado, onde as transações ACID lidam com dados atuais. Contudo, as limitações subjacentes ao modelo relacional tradicional emergem por exemplo em cenários de volumes elevados de informação (100s TiB ou PiB) com *workloads* OLAP onde se pretende consultar de forma *adhoc* informação de histórico;

- **Diversidade** - por razões históricas, as RDBMS tradicionais são proficientes em cenários de gestão transacional de dados de negócio (eg., transações financeiras), contudo atualmente a problemática da gestão de dados estende-se a outros domínios (eg., gestão de catálogos de produtos, motores de recomendação, *big data*, etc.);

- **Consistência e Disponibilidade** - tendo em conta os dois vetores anteriores e o facto de que existem limitações fundamentais inerentes aos sistemas de larga escala, nomeadamente na incapacidade da garantia inequívoca e em simultâneo da consistência da informação e da disponibilidade da aplicação em situação de falha (teorema CAP), as RDBMS tradicionais não são a melhor resposta dado que foram idealizadas para cenários centralizados de consistência atómica (quadrante AC do teorema CAP). A flexibilidade de escolha entre qual o nível de consistência dos dados e disponibilidade da aplicação são primordiais nas soluções contemporâneas de gestão de dados. Tipicamente, as soluções NoSQL são arquitecturalmente desenhadas para lidar com esta problemática.

Aliado aos argumentos apresentados em epígrafe, o fator do elevado custo monetário de uma solução RDBMS tradicional para certos domínios da gestão de informação é facilmente diluído por uma solução NoSQL ajustada ao problema em questão. Este é o ponto essencial que tem relativizado a ideologia "*one size fits all*" face à escolha seletiva de uma base de dados NoSQL para domínios específicos de gestão de dados.

2. TAXONOMIA

Tendo em conta a diversidade de propostas em termos de bases de dados NoSQL, nesta secção são apresentados os principais paradigmas tendo em conta a sua utilidade para o desenvolvimento de soluções e produtos na área das telecomunicações. A apresentação do paradigma em questão foca as principais caraterísticas técnicas, as vantagens e desvantagens de sua utilização, e adicionalmente alguns casos de uso tipicamente cobertos pelo paradigma NoSQL.

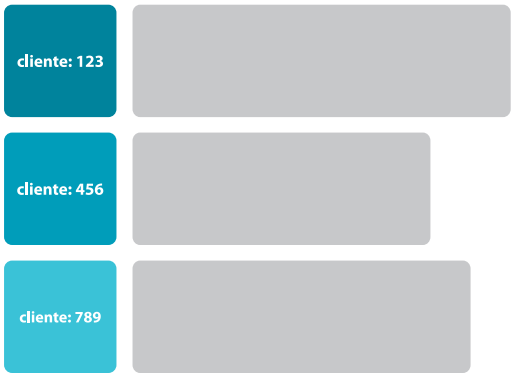
3. *KEY-VALUE*

DESCRIÇÃO

As bases de dados *key-value* são reconhecidamente um dos paradigmas mais simples e mais versáteis no panorama de gestão de dados. Desta forma, assiste-se atualmente a uma oferta extensa de bases de dados deste paradigma e igualmente uma

>>>

panóplia diversa de casos de uso na indústria das telecomunicações. A principal caraterística funcional assenta no armazenamento e endereçamento de dados segundo um modelo de dicionário (*map*) de chaves e respetivos valores (*key-value pairs*).



Os valores persistidos correspondem tipicamente um tipo de dados opaco, desde logo não interpretado pela base de dados e sem pré-definição de *schema* de dados. Desta forma, os valores armazenados podem ter um conjunto variável e mutável de atributos, em conformidade com a dinâmica do negócio.

```
{
"chave": "cliente:123",
"valor": [ opaco ]
}
```

Em complemento, a API exposta pela base de dados *key-value* são caraterizadas pela simplicidade, onde tipicamente são disponibilizadas operações CRUD sobre os dados armazenados em função do parâmetro chave.

```
kstore.create("cliente:123", cliente);
cliente = kvstore.get("cliente:123");
```

Benefícios da utilização de bases de dados *key-value*:

- Possibilidade de armazenamento de dados sem estrutura pré-definida (*schemaless*), em que cada valor pode ter um conjunto dinâmico de atributos;
- A API disponibilizada é simples e compreende normalmente as operações CRUD sobre os pares chave-valor;
- Em termos arquiteturais, as bases de dados *key-value* apresentam poucas limitações ao nível da escalabilidade, seguindo tipicamente uma abordagem *shared nothing* em *hardware* de baixo custo;
- Ao nível da eficiência, a latência de execução das operações CRUD é reduzida e previsível, fruto da simplicidade e da utilização de abordagens de alto desempenho como a de preservar todo o *dataset* em DRAM. O custo monetário por operação CRUD é normalmente bastante reduzido;
- Disponibilizam diversos níveis de garantia de durabilidade dos dados de acordo com a sua criticidade (eg., replicação síncrona ou assíncrona entre réplicas, *commit log* com alterações efetuadas, *snapshot* periódico do *dataset*), permitindo ao utilizador efetuar diversos *tradeoffs* entre durabilidade, disponibilidade e performance no acesso aos dados.

LIMITAÇÕES

Apesar da simplicidade e flexibilidade apresentadas pelas bases de dados *key-value*, naturalmente que a sua proficiência em certos cenários é afetada por algumas limitações inerentes ao paradigma, sendo de salientar:

- Dada a simplicidade da API, tradicionalmente disponibilizam funcionalidades limitadas de pesquisa de dados armazenados, dado que o formato dos dados armazenados é opaco para a base de dados;
- Pouco produtivo em cenários de relacionamentos complexos entre entidades de negócio e na garantia de invariantes na atualização de dados. A lógica de garantia de invariantes e relacionamento de dados é implementada ao nível da aplicação cliente;
- Apesar das garantias de alteração atómica de determinado *key-value pair*, o suporte de transações ACID arbitrárias que envolvam diversos pares chave-valor é limitado ou inexistente.

CASOS DE USO

Conhecidas as principais caraterísticas, vantagens e limitações do paradigma, são seguidamente apresentados os casos de uso mais comuns da sua utilização:

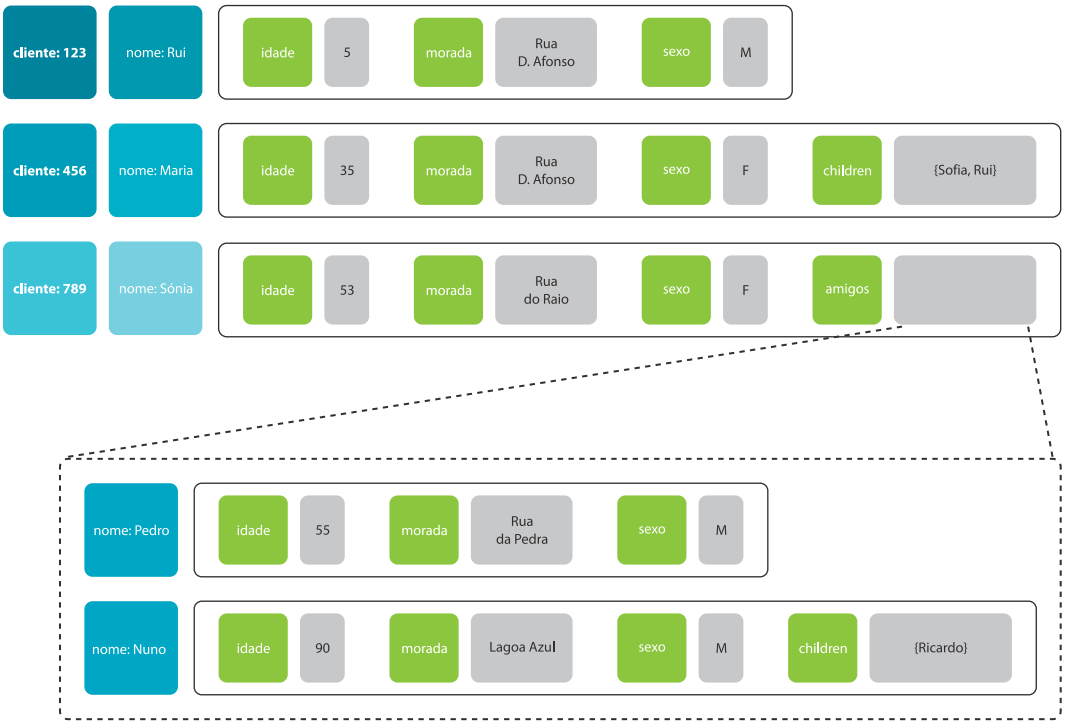
- Armazenamento de *master data* ou dados de referência para consulta ou enriquecimento de informação (eg., cenários de ETL ou *data integration*);
- Gestão de dados com requisitos de baixa latência no acesso (eg., contas de clientes).
- Armazenamento de contadores de negócio com atualização *realtime* (eg., número de execuções do processo de subscrição de produtos).
- cache *sideline* de informação persistida em base de dados tradicional, provendo a solução de baixa latência e menor custo no acesso ao *working set*.

4. DOCUMENT ORIENTED

A tendência para uma crescente interdependência, complexidade e diversidade de dados, originou o

>>>

aparecimento de abordagens de armazenamento de informação em formato semi-estruturado e dinâmico. Nesse sentido, emergiu no seio das bases de dados NoSQL o paradigma de orientação ou documento (*document oriented databases*).Para este facto, contribuíram igualmente as necessidades de acesso aberto e integrado a uma multitude de fontes de dados de sistemas heterogéneos.



O conceito de documento é semelhante ao de registo de dados tradicional, contudo o seu conteúdo é menos rígido e estruturado. Um documento pode ter uma hierarquia dinâmica de atributos de diversos tipos, sendo que, ao invés das bases de dados *key-value*, a estrutura de atributos do documento é de conhecimento da base de dados orientada ao documento. Nesse sentido, o conceito de documento é semelhante ao de um objeto arbitrário tipicamente codificado em formato standard (eg., JSON, XML).

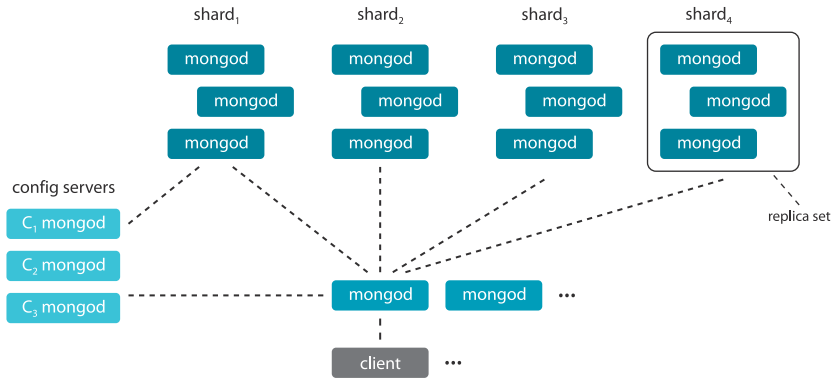
```
{ "cliente: "123",
"nome": "António Pedro", "idade": 35,
"morada": "Rua dos Anjos", "sexo: "M",
"filhos" : ["Sónia", "Pedro"],
"amigos": [{"nome": "Susana", "idade ": 30,
      "morada": "Wall Street", "sexo": "F" },
    { "nome": "Guilherme", "idade": 40,
      "morada": "Rua de São João", "sexo" : "M",
      "filhos" : ["Ricardo", "Rui"],
  }] }
```

Em comparação com uma base de dados relacional tradicional, a *document store* está naturalmente melhor adaptado ao paradigma de modelação orientada aos objetos, não sendo necessário o recurso a tradutores objeto-relacional (ORM).

BENEFÍCIOS

As vantagens típicas da utilização de *document stores* são:

- Proficiência em cenários de dinâmica de atributos associados a determinado objeto (eg., objeto representativo de uma oferta comercial);
- Facilidade no armazenamento e manipulação de dados complexos e polimórficos, tendo a *document store* conhecimento do *schema* do documento armazenado;
- Permite consultas complexas sobre os documentos e seus atributos, com recurso a índices secundários, linguagem de *query* de documentos ou *frameworks* de *map-reduce*;
- Tal como outras soluções NoSQL, é permitida a escolha de diversos níveis de consistência, durabilidade e disponibilidade por tipo de operação a efetuar sobre os documentos (eg., o conceito de *write concern* da MongoDB);
- É comum serem suportadas por uma arquitetura *shared nothing* com enfoque na escalabilidade e eficiência a baixo custo.



LIMITAÇÕES

A utilidade e versatilidade das *document stores* é menos visível e aceitável em determinados cenários, nomeadamente:

- Cenários de relacionamentos complexos entre entidades (eg., motores de recomendação de produtos, modelação e monitorização de redes de equipamentos), dado que não existe o conceito de relação entre documentos. É da responsabilidade da aplicação cliente a materialização do conceito de relações e invariantes entre documentos;
- Cenários de necessidade de transações OLTP sobre um conjunto arbitrário de dados armazenados (eg., Online Charging). De facto, as *document stores* tendem a disponibilizar apenas a atualização atómica de atributos de um documento.

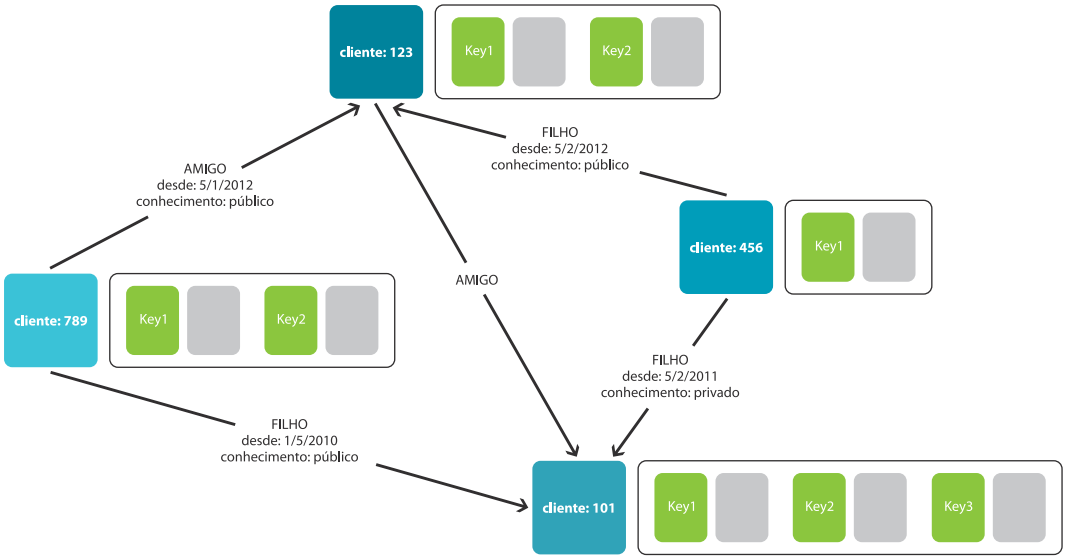
CASOS DE USO

Neste sentido, os casos de uso mais comuns das *document stores* refletem as suas caraterísticas como bases de dados de gestão de informação semi-estruturada (com variância no conjunto de atributos de determinado tipo de documento), nomeadamente:

- Gestão de catálogos de produtos ou serviços;
- Armazenamento de registos de atividade e logs em formato semi-estruturado;
- Base de dados de sistemas de gestão de conteúdos (CMS).

5. GRAPH ORIENTED

Com o advento das redes sociais e da necessidade de representar e relacionar informação de forma arbitrária e não estruturada, emergiu o conceito de bases de dados orientadas ao grafo. Este paradigma NoSQL utiliza as estruturas dos grafos, nomeadamente os vértices e arestas, para representação de informação, permitindo associar propriedades genéricas às estruturas em grafo.



>>>

Em complemento, as bases de dados *graph oriented*, disponibilizam um conjunto de ferramentas e APIs de gestão e pesquisa da informação representada em grafo (eg., caminho mais curto, caminhos alternativos entre vértices, caminhos com determinadas propriedades associadas às arestas). Em alguns casos, é fornecido ao utilizador uma linguagem de *query* orientada ao grafo, aumentando a expressividade no acesso à informação gerida pela base de dados.

```
{ "vertices":[{"cliente 123": {"nome": "Nuno", idade: 23}}
                {"cliente 456: {"nome": "Miguel", idade: 3}}],
  "arestas":[{"origem": "cliente 123",
              "destino": "cliente 456",
              "propriedades": ["amigo", {"desde":"30-01-2013"},
                   {"conhecimento": "público"}
    ]}]
}
```

BENEFÍCIOS

Os itens seguintes sintetizam as principais vantagens da utilização das bases de dados orientadas ao grafo:

- Facilidade de modelação de relações complexas e não estruturadas entre entidades (eg., redes de equipamentos, redes sociais, dependências de software, correlação de falhas em sistemas, etc.);
- Proficiência e baixo custo na pesquisa de dados com relações complexas e arbitrárias (caminhos possíveis com determinada propriedade, caminho mais curto com determinada propriedade, etc.);
- Facilita a construção de ferramentas de visualização de dados relacionados.

LIMITAÇÕES

Adicionalmente, são apresentadas as principais limitações inerentes a estas soluções de gestão de dados:

- Dificuldade inerente de particionamento do dataset (*scale out*) dada a interdependência arbitrária entre os dados armazenados, com tendência para solução arquitetural de único master e diversos *slaves read only*;
- Escalabilidade limitada das operações de criação, remoção e alteração dos elementos do grafo, devido a constrangimentos arquiteturais definidos no ponto anterior;
- Sendo um modelo disruptivo de representação de informação, incorre-se num custo inicial de aprendizagem e domínio da tecnologia.

CASOS DE USO

Os principais casos de uso da tecnologia são:

- Implementação de motores de recomendação de produtos e serviços, sendo a sua implementação facilitada pela disponibilização de linguagens de *query* orientadas ao grafo;

```
START product = node:node_auto_index(name = {product_name})
MATCH (product)<-[:rating5]-(customer)-[:rating5]->(other_stuff)
WHERE other_stuff.type=product.type AND other_stuff.gender=
product.gender
RETURN DISTINCT other_stuff.name
```

- Sistemas de monitoria e alarmística de redes e equipamentos;
- Sistemas de gestão de *workflow* e linhas de montagem;
- Representação das relações entre indivíduos nas redes sociais ou redes colaborativas empresariais.

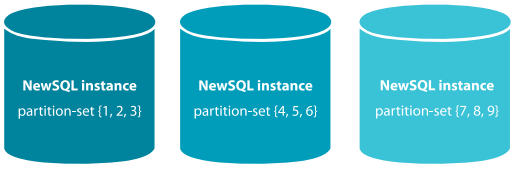
6. NEWSQL

As bases de dados relacionais tradicionais emergiram da necessidade de processamento de transações online (OLTP) sobre dados de negócio, tendo em mente a correção da lógica de negócio subjacente (*correctness*). Contudo, com a evolução tecnológica em curso, as soluções tradicionais de RDBMS são em alguns casos subótimas no aproveitamento dos recursos computacionais disponibilizados. Em concreto no vetor da escalabilidade da solução (capacidade para absorver *workloads* mais expressivos de transações), as estratégias de *scale-up* envolvem um maior custo por transação por unidade de tempo ($/TPS). Neste sentido surgiram as bases de dados NewSQL, como sendo uma classe de bases de dados relacionais com características de performance e escalabilidade de outros paradigmas NoSQL. Tal como as RDBMS tradicionais são desenhados para *workloads* OLTP com garantias transacionais, e suportam a linguagem standard SQL e *stored procedures*. Contudo, possuem caraterísticas arquiteturais que as distinguem em termos de escalabilidade e custo por TPS, nomeadamente,

- O *dataset* é mantido inteiramente em memória DRAM, excluindo desta forma a necessidade de caches ao nível da RDBMS e da persistência de dados orientada ao bloco (eg., árvores B+ tradicionais). As estruturas de dados internas são otimizadas para a persistência em DRAM e para as caches internas dos CPUs disponíveis (*cache conscious*);
- O *dataset* é fragmentado por diversas partições, sendo apenas uma *thread* responsável pela exe-

>>>

cução das transações para determinada partição (*single threaded execution model*). Desta forma, há uma redução dramática dos custos associados ao *context switching, latching* e *locking* inerente nas soluções tradicionais multi-processo. Adicionalmente, é possível desta forma fornecer o nível de isolamento transacional maximo (*serializable*) sem custos adicionais;
- A arquitetura base é *shared-nothing*, sendo possível escalar a solução através da adição de *hardware commodity* (menor custo por TPS).



Em termos de durabilidade de dados e alta-disponibilidade, as bases de dados NewSQL disponibilizam funcionalidades como *snapshot* periódico do *data-set*, replicação síncrona/assíncrona da informação por diversos nós com recuperação automática, e *write-ahead-log* (WAL) com o registo das operações executadas pelas transações.

BENEFÍCIOS

Seguidamente enumeram-se as principais vantagens da utilização de soluções NewSQL face aos sistemas tradicionais RDBMS:

- Maior escalabilidade e menor custo na execução de transações ACID;
- Degradação graciosa em situação de falha com recurso a estratégias de replicação de informação em detrimento de *shared storage*.
- Menos recursos computacionais envolvidos na execução de transações ACID locais a determinado nó, com correspondentes ganhos de eficiência (TPS por instância) e menor custo.

LIMITAÇÕES

As desvantagens mais proeminentes são:

- Custos adicionais de operação e manutenção em cenários com quantidade assinalável de nós;
- Recursos computacionais adicionais na execução de transações ACID que envolvam dados armazenados em diversos nós (eg., recurso ao protocolo *two-phase commit* durante a execução da transação);
- Maturidade moderada de algumas soluções NewSQL, sem provas dadas em cenários críticos.

CASOS DE USO

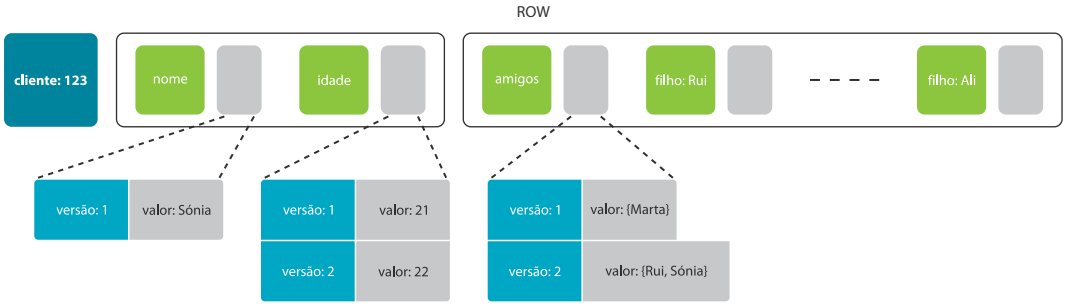
Os casos de uso mais comuns são designadamente os cenários OLTP com *workloads* intensivos de transações ACID (TPS elevados) e sensíveis á latência de execução, como por exemplo:

- Transferências financeiras e de mercados de capitais;
- Plataformas de jogo online ou colaborativo;
- Cobrança *online* de utilização de produtos ou serviços;
- Cenários de processamento de ordens de compra no retalho *online* ou tradicional, envolvendo grandes quantidades de transações ACID.

7. *BIGTABLES*

Para além da diversidade e dinâmica da estrutura dos dados a gerir (dados semi-estruturados), outro vetor importante é o volume crescente da informação a armazenar e processar. Este aspeto é especialmente relevante em sistemas OLAP (*online analytical processing*) onde são efetuadas consultas sobre uma quantidade assinalável de informação. Nesse sentido, desenvolveu-se o conceito das *bigtables*, como um novo paradigma de gestão de dados adaptado à nova realidade do *Big Data*.

Tecnicamente, uma solução *bigtable* faz a gestão de tabelas de células de dados, em que uma tabela é vista como um dicionário (*map*) ordenado pela chave da linha. O dicionário é multidimencional e *sparse* (podendo ter células sem dados para determinada coluna). Adicionalmente, suportam nativamente a gestão de várias versões das células de dados.

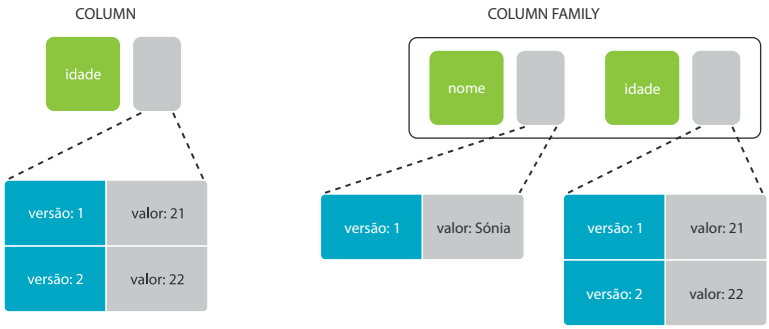


>>>

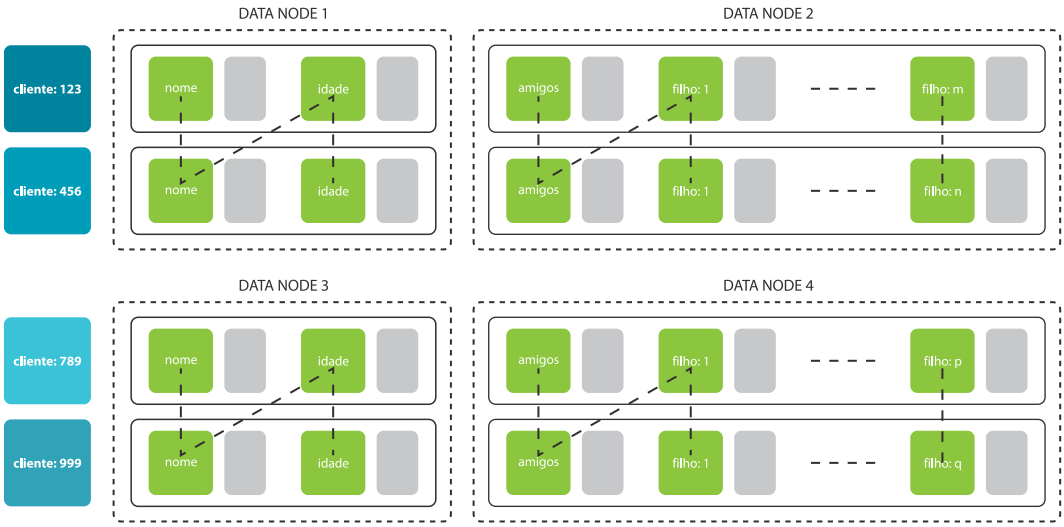
Tendo como exemplo a figura em epígrafe, o endereçamento de uma célula de informação seria descrito como:

{chave: "cliente:123", coluna: "idade", versão: "2"} ⟶ valor: "22"

Como se pode constatar, a coluna é utilizada para o armazenamento de células de dados diferenciadas por versão. Adicionalmente, é possível organizar as colunas relacionadas em famílias de colunas, sendo esta a unidade de agregação de dados para armazenamento.



O armazenamento dos dados tabulares organizados em famílias de colunas é efetuado em secções ordenadas de colunas (*column oriented*) e de forma distribuída por diversos nós, conforme na figura seguinte.



É de realçar que, dado o esquema de armazenamento ser orientado à coluna, os valores de determinada coluna encontram-se armazenados de forma ordenada pela chave do registo. Desta forma, o cálculo de agregados ou consulta filtrada da informação por range de chaves é sequencial e paralela entre diversos nós de dados.

BENEFÍCIOS

Os principais benefícios das BigTables são:

- Adaptadas a cenários de *Big Data*, sendo bastante proficiente na computação de agregados que envolvam apenas um sub-conjunto de colunas, tipicamente pertencentes á mesma família de colunas (*data locality*);
- Com o armazenamento *column oriented*, há mais oportunidades de otimização de espaço de armazenamento, por exemplo com o recurso a estratégias de compressão *column-aware* (eg., *run-length encoding*);
- Capacidades de *scale out* com agregação de diversos nós de armazenamento e processamento paralelo;
- Proficiente em cenários OLAP (*write once & read many*).

LIMITAÇÕES

As limitações mais significativas compreendem:

- Menor eficiência em cenários OLTP, dado que a informação da linha encontra-se dispersa em termos de armazenamento;

>>>

- O peso associado aos mecanismos internos de *garbage collection* de versões de dados antigas;
- Alguma complexidade na definição de um modelo de dados eficiente, com forte dependência das *queries* a executar e arquitetura de *storage*.

CASOS DE USO

Os casos de uso mais comuns deste paradigma de gestão de dados são:

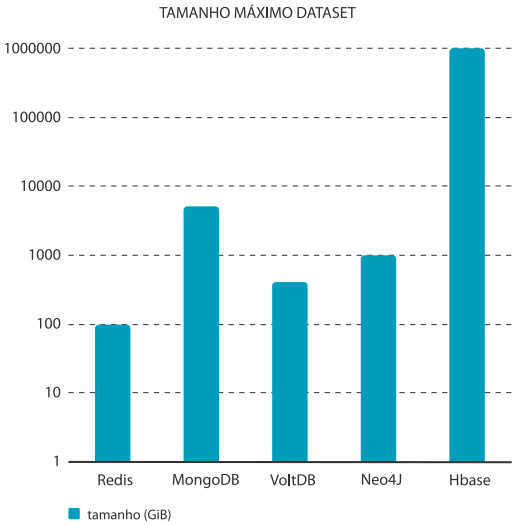
- Armazenamento de grandes quantidades de registos de atividade, com posterior extração de indicadores, contadores, etc.;
- Suporte á implementação de sistemas de análise de dados e extração de conhecimento;
- Armazenamento temporário de grandes quantidades de dados numa solução de *data hub*.

8. ANÁLISE QUANTITATIVA

Nesta secção são apresentados alguns indicadores quantitativos testados referentes a soluções NoSQL, tendo em conta um cenário de negócio que envolve a compra online de produtos. A análise comprometeu a escolha de bases de dados NoSQL de diversos paradigmas, nomeadamente:

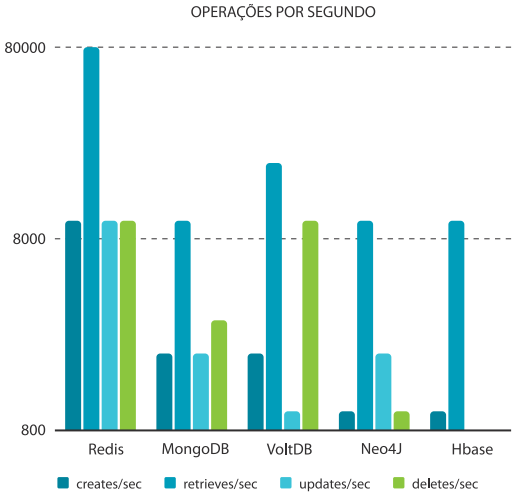
BASE DE DADOS	PARADIGMA
Redis	key-value
MongoDB	document oriented
VoltDB	NewSQL
Neo4J	graph oriented
HBase	BigTable

No cenário concreto, foram avaliadas quantidades em termos do tamanho máximo do *dataset* suportado pelas soluções NoSQL, conforme demonstrado na figura seguinte (note: escala logarítmica):



Como seria espectável, a base de dados HBase como um exemplo de uma BigTable é a que suporta a gestão do *dataset* de dimensão maior.

Adicionalmente, foram tidas em consideração métricas de desempenho das operações mais frequentes. A figura seguinte apresenta a métrica de *thoughput* das operações CRUD disponibilizadas pelas APIs das bases de dados testadas (note: escala logarítmica):



É de realçar o desempenho assinalável da *key-value store* Redis na execução de operações CRUD.

REFERÊNCIAS

Bases de dados NoSQL – Coordenação Tecnológica PTIN (http://wiki.ptin.corppt.com/display/CT/Bases+de+dados+NoSQL)

CVS DOS AUTORES

Manuel Francisco Gonçalves, encontra-se concluir o Mestrado em Engenharia de Sistemas e Informática na Universidade do Minho. É colaborador da PT Inovação desde 2012, tendo participado no desenvolvimento e investigação de sistemas Big Data da PT Inovação. Atualmente é membro do grupo de coordenação tecnológica do departamento CTE.

Mário Moreira obteve o mestrado em Ciências da Computação pela Universidade do Minho em 1995 e a Licenciatura em Engenharia de Sistemas e Informática pela Universidade do Minho em 1993. Ingressou no Centro de Estudos de Telecomunicações (futura PT Inovação), ainda durante a licenciatura em 1993 onde esteve a desenvolver vários sistemas periciais baseados em técnicas de Inteligência Artificial. Esteve inicialmente envolvido em vários projetos europeus de investigação ACTS e Eurescom. Desempenhou funções na área de serviços de gestão de redes, tendo colaborado com a equipa da Portugal Telecom no âmbito da Expo98, e estando envolvido no desenvolvimento de vários produtos da PT Inovação nesta área. Em 2001 ficou responsável pela Unidade de Plataformas, Serviços e Aplicações para Redes Móveis, onde esteve envolvido no desenvolvimento de várias plataformas de Unified Messaging, SMS-C, Localização e Portais Móveis. Atualmente é Gestor da Divisão de Tecnologias e Desenvolvimento na direção de Coordenação Tecnológica e Inovação Exploratória.

Miguel Biscaia, Licenciado em Engenharia de Sistemas e Informática pela Universidade do Minho em 1999. É colaborador da PT Inovação desde 2002, tendo participado no desenvolvimento de sistemas OSS e BSS da PT Inovação. Atualmente é membro do grupo de consultadoria e coordenação tecnológica do departamento CTE.

04 Cloud Platform-as-a-Service



DAVID CUNHA PEDRO NEVES PEDRO SOUSA

PALAVRAS CHAVE
Cloud Computing, Service Broker, Platform-as-a-Service, APIs, Interoperabilidade

O *Cloud Computing* tem emergido como sendo um novo paradigma para entrega de serviços através da Internet. Neste mercado em expansão, o serviço de PaaS (*Platform-as-a-Service*) tem sido objeto de grande interesse por parte das mais variadas organizações permitindo o fácil *deployment* de aplicações sem necessidade de uma infra-estrutura dedicada, instalação de dependências ou configuração de servidores. No entanto, cada fornecedor de soluções PaaS acaba por gerar um *lock-in* do utilizador às suas características proprietárias, tecnologias ou APIs (*Application Programming Interfaces*). Além disso, dando como garantida a conectividade até aos clientes, a rede de operadores como seja o caso da PT (Portugal Telecom) acaba por servir apenas de *dumb-pipe* entre o fornecedor e os seus clientes.

Resumidamente, os principais objetivos deste artigo são apresentar os resultados do projeto CSB (*Cloud Service Broker*), realizado em colaboração com o Instituto de Telecomunicações de Aveiro, que tem como intuito resolver as questões de interoperabilidade entre fornecedores de PaaS; e igualmente apresentar o trabalho desenvolvido no projecto FP7 Cloud4SOA no qual a PTIN fez parte do consórcio.

1. INTRODUÇÃO

Além de um *buzzword* como o próprio termo *web* o é, *Cloud Computing* é uma evolução de vários paradigmas tecnológicos das últimas décadas transformando o sonho do *computing-as-a-utility* numa realidade. O potencial de tal modelo preconiza um grande impacto na indústria das TIs onde os recursos computacionais e o *software* são entregues aos utilizadores finais através de um paradigma *pay-per-use* [1].

Neste mercado em expansão, o serviço de PaaS tem sido objecto de grande interesse por parte das mais variadas organizações permitindo o fácil *deployment* de aplicações sem necessidade de infra-estruturas dedicadas, instalação de dependências ou configuração de servidores. Nos últimos anos, sensivelmente desde de 2009, surgiram diversos fornecedores e *startups* de soluções PaaS que competem entre si arduamente focando-se nos aspetos inovadores que os possam diferenciar face ao utilizador. Portanto, cada fornecedor tem intrinsecamente associadas diferentes linguagens de programação, *frameworks*, base de dados, taxonomias, modelos de negócio, ferramentas de desenvolvimento e APIs (*Application Programming Interfaces*) para interação por parte dos seus clientes. Por um lado, esta mescla de fornecedores favorece o processo de seleção por parte do utilizador. Mas por outro, surge a possibilidade de existir um *lock-in* do cliente a um fornecedor específico e às suas características proprietárias.

A Portugal Telecom Inovação (PTIN) pretende fornecer os alicerces para a entrada no mercado de *cloud* de uma plataforma mediadora entre os utilizadores e os fornecedores de PaaS de forma a posicionar-se firmemente na cadeia de valor deste mercado. Assim sendo, a PTIN teria a capacidade de oferecer aos clientes a possibilidade de operar sobre várias plataformas de forma centralizada e abstraindo os seus utilizadores das diferenças intrínsecas na interação com cada fornecedor suportado.

2. *PLATFORM-AS-A-SERVICE*

Até este momento, o IaaS (*Infrastructure-as-a-Service*) continua a ser o modelo de serviço *cloud* mais utilizado e com mais sucesso na área. Todavia o PaaS, sendo orientado ao desenvolvimento de aplicações, possui o potencial para abstrair as organizações de todos os processos de manutenção e configuração de servidores, bem como de instalação de dependências [2]. Um fornecedor de PaaS oferece um ambiente integrado simples e intuitivo para desenvolver, testar, monitorizar e hospedar aplicações *web* e respetivas bases de dados. Apesar do mercado de PaaS ainda se encontrar numa fase precoce, a Forrester[1] estimou que em 2016 o volume gerado devido ao mercado de PaaS poderia atingir os 15,2 mil milhões de dólares [3]. No seio de uma organização, o impacto destes ambientes, tanto a nível técnico e económico, é enorme. Os vários profissionais das TIs, nomeadamente programadores, administradores de sistemas e até gestores empresariais, são abrangidos por esta revolução. Em poucos dias um simples protótipo de um serviço pode ser lançado para produção sem existir um grande investimento inicial por parte da organização detentora. Subitamente, a inovação cessa de estar só no horizonte de quem possui avultados recursos económicos para transformá-la num produto comercializável.

No mercado surgem diversos fornecedores de PaaS cada um oferecendo um serviço diferenciado

[1] http://www.forrester.com/home/

para o utilizador. Além dos titãs Amazon, Google e Windows, emergem ofertas especializadas em linguagens de programação específicas, caso do CloudBees, e igualmente soluções *open-source*, nomeadamente o OpenShift da Red Hat e o Cloud Foundry da VMware. O Heroku da Salesforce.com surge como um dos principais fornecedores de PaaS especializado em linguagens *open-source* tais como Ruby e Python, e suportando Git como ferramenta de controlo de revisão. Dessa forma será feita uma síntese desses fornecedores.

2.1 CLOUDBEES

O CloudBees [4] foi fundado em 2010 e é um PaaS inteiramente direcionado ao desenvolvimento de aplicações Java. Portanto, suporta qualquer linguagem de programação e *framework* que executem sobre uma JVM (Java Virtual Machine), nomeadamente Java, JRuby, Scala, Play, Grails, Spring, etc. O CloudBees fornece dois serviços, sendo um deles orientado ao desenvolvimento e teste de aplicações, e outro orientado apenas ao *deployment* e execução das mesmas. Através do *DEV@Cloud*, o utilizador tem acesso a ferramentas como Jenkins, Maven, Sonar e repositórios Git e SVN de modo a controlar todo o ciclo de vida da aplicação. Por sua vez através do *RUN@Cloud*, o cliente poderá efetuar rapidamente o *deployment* de uma aplicação, já criada previamente, adquirindo acesso a todas as funcionalidades de gestão, monitorização, *logging* e escalabilidade.

2.2 CLOUD FOUNDRY

O Cloud Foundry [5] é um PaaS que possui uma abordagem dissemelhante em comparação a quase todas as restantes plataformas do mercado devido a ser uma plataforma completamente *open-source* com licença Apache v2. Lançado em 2011 pela VMware, o Cloud Foundry prima por suportar diversos ambientes de execução (e.g. Java, Scala, Ruby, Node.js, etc.), *frameworks* (e.g. Spring, Sinatra, Rails etc.) e base de dados (e.g. MySQL, MongoDB, Redis, PostgreSQL), sem estar fixo a uma única infra-estrutura. O utilizador tem assim a oportunidade de alterar o código fonte do próprio PaaS e assentá-lo sobre qualquer serviço de infra-estrutura ao seu dispor, seja público ou privado. Existem diversas soluções baseadas no Cloud Foundry como o AppFog, Stackato e a extensão .NET: IronFoundry.

2.3 HEROKU

O Heroku [6] surgiu em 2007 como sendo um PaaS orientado exclusivamente ao desenvolvimento de aplicações Ruby. Após ser adquirido pela Salesforce.com em 2010, o Heroku evoluiu suportando mais tecnologias e tornando-se um dos PaaS mais utilizados e conhecidos do mercado. De momento, é estimado que suporta mais de 1,5 milhões de aplicações de todo o tipo de linguagens de programação desde de Java, Scala, Python, Ruby, PHP ou Node.js., com base de dados PostgreSQL ou Redis. O Heroku possui um amplo catálogo de *addons* que permite aos utilizadores adicionarem diversos tipos de base de dados SQL e NoSQL, serviços de monitorização, *logging*, faturação, teste, etc. O Git tem-se tornado um de facto no mercado de PaaS sendo cada vez mais utilizado como ferramenta de *deployment* e controlo de revisão de código fonte em diversos PaaS, não sendo exceção no Heroku.

3. *CLOUD SERVICE BROKER*

Cada aplicação possui dependências específicas que podem ser tão distintas como a mesma ser desenvolvida em Java ou Python até necessitar de uma base de dados SQL ou NoSQL. Para um utilizador que pretenda portar as suas aplicações para um ambiente de *cloud*, ele precisará de examinar as diversas ofertas existentes no mercado, conhecer as tecnologias suportadas, estudar e testar os interfaces cliente, as APIs fornecidas, etc. Dessa forma surgiu a plataforma CSB (*Cloud Service Broker*), que foi desenvolvida no âmbito de um projeto do Instituto de Telecomunicações de Aveiro suportado pela PTIN, com o intuito de facilitar a descoberta e interacção com diversos fornecedores de PaaS através de um único interface.

A arquitectura definida inclui além do CSB, o *PaaSManager*, que será descrito mais à frente, e as interfaces cliente: *web*, CLI e Git. Através das interfaces *web* e CLI, o utilizador tem acesso às operações dos serviços de gestão e de informação suportados pelo *PaaSManager*. No momento da criação de uma aplicação é preenchido o formulário que constitui o perfil tecnológico da aplicação, nomeadamente, os *runtimes, frameworks* e *services* (base de dados). Desta forma, o CSB invoca o *PaaSManager* para devolver as tecnologias suportadas por cada plataforma do ecossistema, comparando com a informação inserida no perfil. Por fim, é devolvida uma lista ordenada com os fornecedores recomendados. Nos diversos fluxos de operações teve-se em atenção que do ponto de vista de utilizador a interacção com esta arquitectura não se verificasse mais complexa do que interagir directamente com um fornecedor específico.

Na Figura 1 é apresentada a arquitetura do *Cloud Service Broker* bem como a sua integração com os restantes módulos desenvolvidos.

>>>

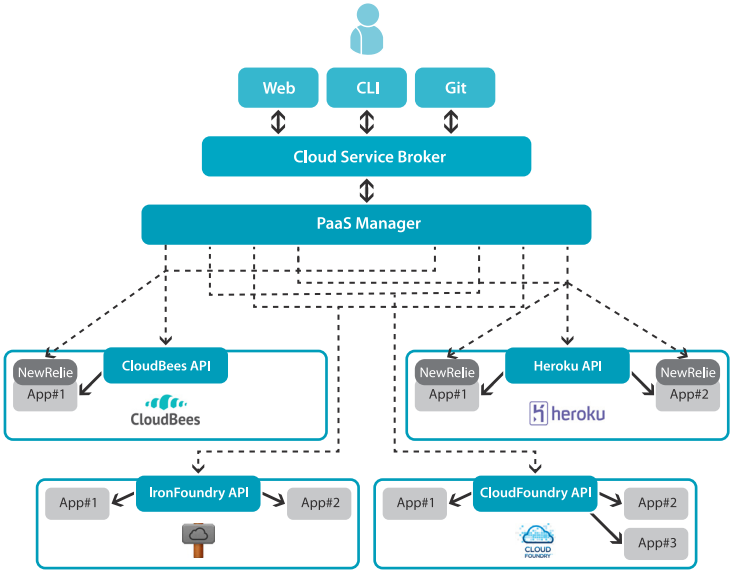


Figura 1. Arquitetura *Cloud Service Broker*

3.1 RECOMENDADOR *CLOUD SERVICE BROKER*

O *Cloud Service Broker* visa efetuar a recomendação baseada em regras definidas e analisadas pelo seu motor de regras. As suas características principais são dar assistência no processo de recomendação e fornecer todas as operações suportadas através de um único interface. Portanto, o utilizador tem assim acesso a todo o ciclo de vida de uma aplicação hospedada no PaaS recomendado pelo CSB. Além dos requisitos técnicos de uma aplicação, o algoritmo de recomendação poderá ser estendido suportando diversas exigências. São alguns exemplos a geração de notificações quando certos patamares nos valores de monitorização forem ultrapassados, de forma a respeitar os contratos definidos pelo próprio utilizador; ou então a recomendação por base nos modelos de negócio que se verifiquem mais adequados para o próprio detentor de uma aplicação.

Esta plataforma foi desenvolvida em Java, o algoritmo de recomendação em Prolog, e foi utilizada uma base de dados MySQL para registo de utilizadores e manutenção de estado.

3.2 *PAASMANAGER*

O *PaaSManager* é uma camada de abstração que visa unificar os processos de gestão e aquisição de informação de aplicações criadas através de diversos PaaS de modo a combater o *lock-in* existente no mercado. Inicialmente foram seleccionadas algumas plataformas que surgem no mercado com o intuito de implementar a solução que agrega as diferentes ofertas de forma transparente para o utilizador. Cada fornecedor de PaaS disponibiliza uma API. Através dessa interface são expostas diversas funcionalidades que permitem criar, gerir e obter informação sobre aplicações e bases de dados. Porém, do leque de métodos disponibilizados, foi fundamental eleger as funcionalidades idênticas ou semelhantes que são partilhadas entre as várias APIs analisadas. Apenas dessa forma seria possível obter a interoperabilidade necessária para o utilizador interagir de igual forma com diferentes PaaS que possuem diferentes APIs. Das plataformas investigadas foram selecionados para o ecossistema 4 fornecedores: CloudBees, Cloud Foundry, Iron Foundry e Heroku. Todos os fornecedores possuem diferentes APIs, ferramentas de monitorização e de *deployment*, em exceção do Cloud Foundry e do Iron Foundry que partilham uma implementação de API idêntica, porém suportando tecnologias discrepantes. Desta forma, o *PaaSManager* pode ser integrado com o recomendador *Cloud Service Broker* através de uma única interface que abstrai os vários fornecedores suportados.

No desenho da arquitetura do *PaaSManager* teve-se em atenção a modularidade preponderante que permite que todo o sistema continue em completo funcionamento mesmo que alguma API de um fornecedor ou uma API de monitorização não esteja a operar corretamente. Consequentemente, cada API foi implementada por diferentes módulos de software e gerida por entidades únicas orientadas aos grupos de serviços de gestão e de informação. Por fim, uma interface RESTFul expõe as várias operações especificadas que podem ser invocadas por um simples cliente HTTP.

A Figura 2 ilustra a arquitetura do *PaaSManager*, os diversos módulos que foram especificados e a sua integração com as diversas APIs dos fornecedores de PaaS.

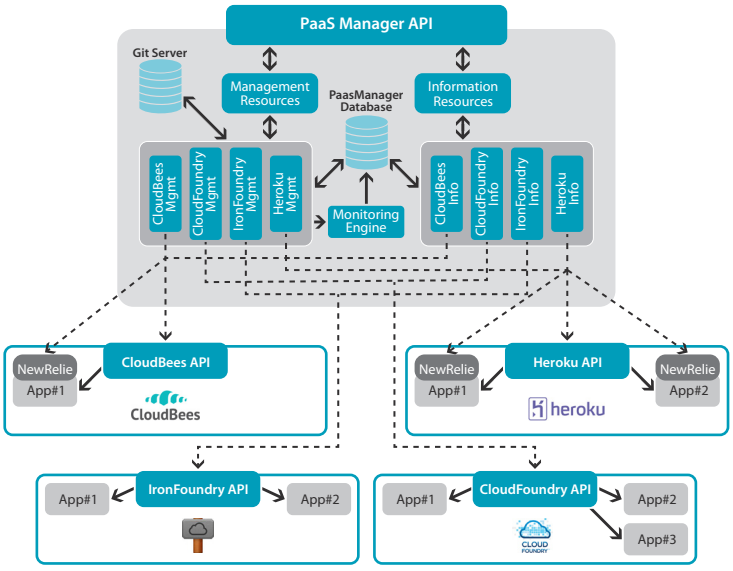


Figura 2. Arquitetura *PaaSManager*

3.2.1 Principais Operações

Nesta secção são apresentadas algumas das operações mais fundamentais que são suportadas pelo *PaaSManager*.

- ***Deployment* da Aplicação**

O *deployment* do código fonte foi unificado para que o utilizador não necessite de se preocupar com as diferentes ferramentas oferecidas pelos fornecedores. A Figura 3 ilustra o processo ocorrido durante a operação de *deployment*. O pedido recebido pela API é instantaneamente enviado ao módulo de gestão, *Management Resources*, que realiza uma busca na base de dados central com o objectivo de adquirir a identificação da plataforma onde se encontra criada a aplicação (passo prévio). Através do resultado obtido é invocado o *adapter* do PaaS pretendido, que por sua vez, executa os comandos *git-add* e *git-commit* no repositório da aplicação que se encontra no servidor Git central. Após este processo, é efetuado o *deployment* na plataforma conforme o paradigma associado. No caso do CloudBees, é feita uma pesquisa no repositório até ser encontrado o *web archive* da aplicação (.war). Para o Cloud Foundry e Iron Foundry é realizado o mesmo processo para aplicações baseadas em Java, porém, para as restantes é enviado todo o código fonte. Para o Heroku é executado um comando *git-push* para o repositório remoto criado exclusivamente para a aplicação na plataforma Heroku.

Numa situação de sucesso, é iniciado o motor de monitorização com o objetivo de recolher estatísticas em tempo-real sobre a aplicação, armazenando-as na base de dados central. Para finalizar o processo, é devolvida uma resposta em XML ou JSON ao utilizador. A estratégia de funcionamento adotada para o motor de monitorização será posteriormente analisada com mais detalhe.

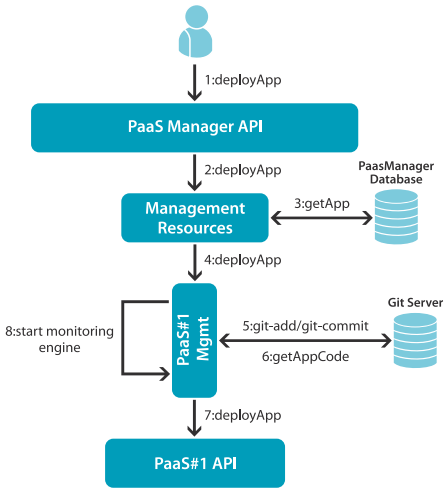


Figura 3. *PaaSManager - Deploy App*

- **Adquirir Estado da Aplicação**

O acesso à informação sobre o estado físico de uma aplicação a executar em uma plataforma é de facto fundamental para o utilizador. A Figura 4 ilustra os diversos processos que ocorrem durante esta operação. De forma a dar uma simples e intuitiva visão sobre o estado de qualquer aplicação, foram especificados 4 estados: *running, stopped, crashed* e *unknown*. No processo de obtenção de estado, o pedido efetuado à API do *PaaSManager* é instantaneamente enviado ao módulo de informação, Informação *Resources*, que direciona a mensagem

para o *adapter* do PaaS pretendido. Por sua vez, o *adapter* invoca o método da API da plataforma associada. Conforme a informação de estado devolvida, a mesma é mapeada para um dos 4 estados definidos com o intuito de unificar a resposta XML ou JSON para o utilizador.

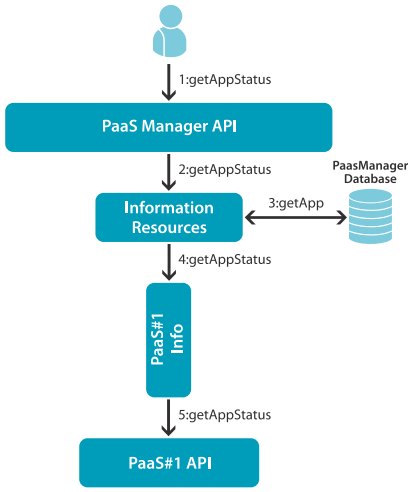


Figura 4. *PaaSManager* - Adquirir Estado da Aplicação

- **Monitorização**

A monitorização em tempo-real de aplicações permite fornecer ao utilizador a informação necessária sobre o comportamento das mesmas nos diferentes PaaS suportados. Até ao momento não existe um consenso nem nenhum modelo normalizado de monitorização partilhado entre os maiores fornecedores. Nos últimos anos, os estudos efetuados nesta área tentam definir *frameworks* de monitorização como também um modelo de métricas que se verifique apropriado para uma gestão eficiente de aplicações na *cloud* [7], [8]. Uma normalização do modelo de monitorização faria com que os SLAs definidos pelos fornecedores fossem comparáveis e o utilizador pudesse testar a mesma aplicação em 2 ambientes heterogéneos com o objetivo de obter informação sobre o comportamento dos ambientes disponibilizados. No entanto, ao nível do serviço de PaaS, cada fornecedor possui diferentes métricas e formas de monitorizar as aplicações sendo que grande parte da informação relacionada com os recursos de infraestrutura é abstraída para o utilizador. Como foi analisado, o CloudBees e o Heroku possuem parceria com APMs (*Application Performance Monitor*), nomeadamente, o New Relic que permite a monitorização de aplicações através da instalação de agentes nas instâncias onde as próprias executam. Por outro lado, o Cloud Foundry e Iron Foundry devolvem outro tipo de estatísticas diretamente através da API nativa.

Neste contexto, foi especificado um módulo na arquitetura do PaaS Manager, denominado por *Monitoring Engine* (ver na Figura 2), que efetua a recolha em tempo-real das métricas expostas pelas diferentes plataformas do ecossistema. A informação obtida é depois armazenada na base de dados central para ser devolvida ao utilizador através de pedidos à API do *PaaSManager*. O processo é definido por uma recolha síncrona efetuada todos os minutos à API do New Relic ou à API nativa, conforme o PaaS onde a aplicação se encontra hospedada. A Figura 5 ilustra os diversos passos que são executados durante o processo de monitorização.

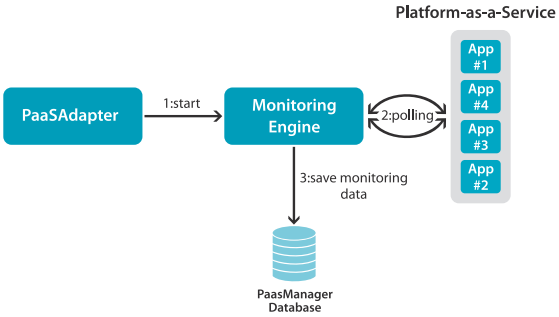


Figura 5. *PaaSManager - Monitoring Engine*

Toda a implementação do *PaaSManager* foi realizada em Java recorrendo à plataforma JavaEE6 (Java Enterprise Edition 6) devido à ótima integração com EJBs (Enterprise Java Beans), *web services* e camadas de persistência. O servidor aplicacional JBoss 7.1.1 foi o servidor selecionado para suportar o *PaaSManager* no ambiente de desenvolvimento e teste.

4. CLOUD4SOA

Context-as-a-Service traz a oportunidade de usar a informação gerada pelos clientes num ambiente móvel para selecionar e entregar conteúdo mais adequado para os mesmos (por exemplo, vídeos, fotos, etc.). Estes meta-dados podem caracterizar qualquer situação do cliente, tal como a sua posição, género, como simultaneamente interesses de musicais ou sociais. No âmbito do projeto FP7 C-CAST [9] que finalizou em 2010, a PTIN desenvolveu uma *framework* composta de diversos serviços que reagem a informação de contexto de um utilizador. Neste contexto, a proliferação de tais serviços móveis pode, em alguns casos, gerar uma sobrecarga na infra-estrutura subjacente se a mesma não for escalável.

A PTIN no projeto Cloud4SOA teve como grande objetivo migrar os vários serviços da *framework* de contexto para um ambiente *cloud*, nomeadamente, PaaS. Essa migração envolveu um extenso estu-

do sobre as alterações a realizar nas aplicações. A arquitetura, apresentada na Figura 6, é composta por vários módulos que permitem recolher e processar informação relativa ao terminal móvel do cliente. Como componente central e fundamental da *framework* surge um servidor XMPP denominado por Context Broker. Este servidor tem um papel ativo nos processos efetuados pelos 3 serviços de contexto: *Context Enabler, Group Manager Enabler* e *Content Selection Enabler*.

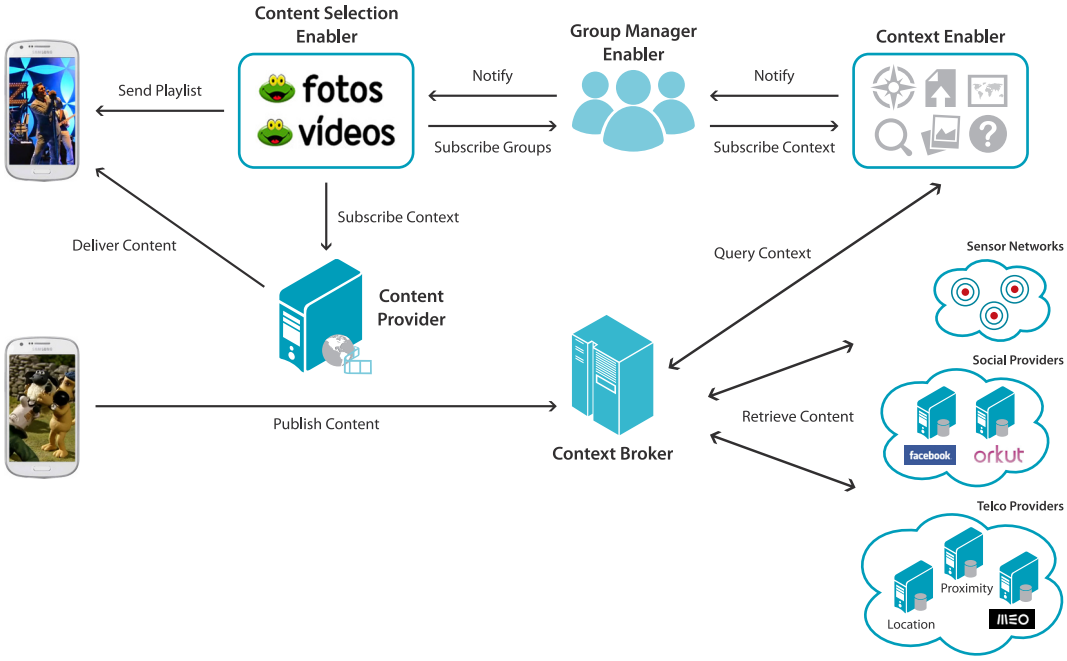


Figura 6. Cloud4SOA - Arquitetura *Framework* Contexto

Dos vários componentes foram migrados os seguintes 3 serviços:

- ***Context Enabler***: Este serviço foi desenhado durante o projeto Cloud4SOA. É um *web service* SOAP com o propósito de servir de interface para o servidor XMPP e permitir a reutilização segundo os conceitos SOA.
- ***Group Manager Enabler:*** Este serviço é responsável por estabelecer grupos de utilizadores baseado em condições pré-configuradas.
- ***Content Selection Enabler***: Este serviço tem por base um motor de regras que selecciona conteúdo (fotos, vídeos, etc.) mais adequados para os grupos de utilizadores.

Todas as aplicações foram preparadas para executarem sobre o servidor Apache Tomcat devido a ser um dos servidores Java mais suportados pelos fornecedores de PaaS existentes no mercado. Para o processo de migração foi utilizado o sistema desenvolvido pelo restante consórcio do Cloud4SOA tendo sido realizados alguns testes que provaram uma boa escalabilidade dos serviços como igualmente custos mais reduzidos para instanciar e manter a mesma infra-estrutura que existia *on-premise*.

5. ENSAIOS E AVALIAÇÃO

Nesta seção são descritos alguns dos ensaios e avaliações realizadas no projeto *Cloud Service Broker*.

5.1 INTERFACE *WEB*

O portal *web* do *Cloud Service Broker* foi desenvolvido em *Ruby on Rails* tendo sido utilizado *HTML, CSS, JQuery* e a *framework Twitter Bootstrap* para o *frontend*.

Na Figura 7 é possível observar a página inicial do portal onde o cliente pode iniciar o processo de recomendação em *New App*.

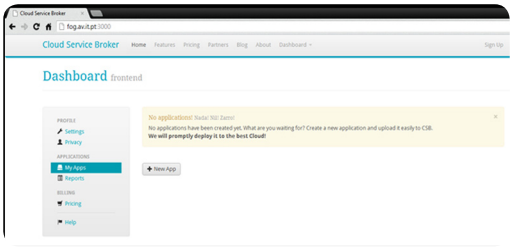


Figura 7. *Cloud Service Broker - Home Page*

De seguida, na Figura 8, é apresentado ao cliente um formulário onde o mesmo pode preencher as características técnicas da aplicação que deseja migrar para a *Cloud*. Como resultado é apresentada uma lista de fornecedores mais adequados.

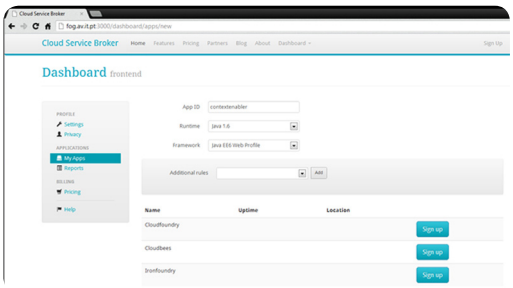


Figura 8. *Cloud Service Broker* - Recomendação

Após seleccionar o fornecedor mais adequado, a aplicação foi criada com sucesso. O processo prossegue com o *deployment* do código fonte através da ferramenta Git e para o respectivo repositório remoto.

Desta forma, a aplicação encontra-se *online* tal como demonstrado na Figura 9.

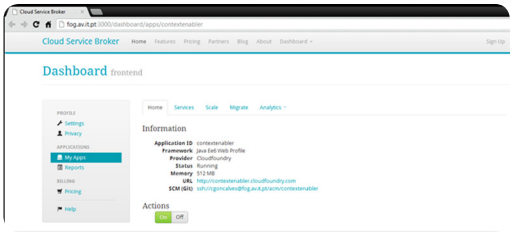


Figura 9. *Cloud Service Broker - App deployed*

Por fim, como se pode observar na Figura 10, o utilizador tem acesso às várias métricas recolhidas no fornecedor.

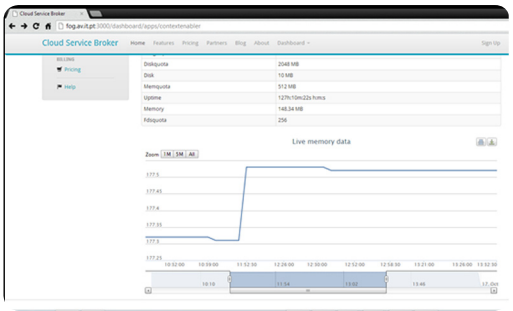


Figura 10. *Cloud Service Broker* - Monitorização

5.2 *PAASMANAGER*

Para testar a solução *PaaSManager* foram selecionados alguns dos métodos fundamentais de forma a serem avaliados diferentes aspetos e depurado onde seria possível proceder a otimizações. Através da ferramenta Apache JMeter foram realizados alguns testes de carga simulando o acesso de diversos clientes em simultâneo. Os valores obtidos serviram de referência para compreender o comportamento que o sistema detém em operações essenciais. Por um lado, de forma a se observar o *overhead* que é acrescido pelo *PaaSManager* em cada pedido efetuado, e por outro, de forma a se analisar a escalabilidade da solução com diferentes números de utilizadores. Para cada uma das operações selecionadas foram realizados ensaios com 10 e 30 utilizadores em simultâneo, obtendo-se a média e o desvio padrão associado a cada operação efetuada nas diferentes plataformas que constituem o ecossistema. Esse número de utilizadores foi escolhido devido ao *PaasManager* utilizar apenas uma conta para cada PaaS. Assim sendo, um grande número de pedidos poderia se refletir num bloqueamento da conta ou num efeito de *throttling* por parte dos fornecedores. É necessário realçar que o objetivo desta análise não foi comparar diretamente a eficiência de cada fornecedor mas sim o desempenho da arquitetura desenvolvida. Como exemplo ilustrativo apresenta-se de seguida alguns resultados relacionados com a operação de aquisição de estado da aplicação.

- **Adquirir Estado da Aplicação**

Nesta operação foram obtidos, com um total de 30 utilizadores em simultâneo, os resultados apresentados na Figura 11. A série *PaaSManager*, que representa o tempo consumido apenas no processamento interno do pedido, apresenta valores aproximados de tempo de resposta entre os 60 e os 800 ms. Esta série apenas inclui o processo de aquisição da identificação da plataforma até ser invocado o respetivo *adapter* de PaaS. Na série *PaaSManager+PaaS API*, que abrange o tempo consumido em todo o processo da operação incluindo a solicitação à API da plataforma em questão, os diversos fornecedores registaram os valores entre os 1200 e 1600 ms. No caso dos *adapters* do Cloud Foundry e Iron Foundry existe uma verificação nos *logs* de cada instância onde a aplicação executa com o intuito de detetar erros de funcionamento, explicando assim a maior contribuição do *PaaSManager* nos tempos de resposta obtidos.

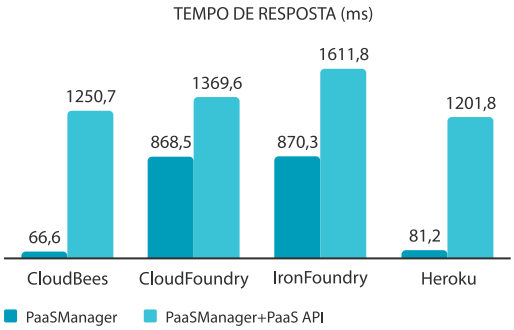


Figura 10. *PaaSManager* - Adquirir Estado App (30 Utilizadores)

6. CONCLUSÕES

O *Cloud Service Broker* e o Cloud4SOA foram projetos pioneiros na PTIN na área de PaaS. Os resultados trouxeram um considerável *know-how* sobre estes ambientes de *cloud*, originando também novas questões e desafios para futuros trabalhos.

No projeto Cloud4SOA foram migradas para um ambiente *cloud* algumas das aplicações que fazem parte de uma *framework* de contexto já existente. Para realizar essa migração com sucesso, algumas das aplicações foram modificadas sendo que os resultados obtidos mostraram as valências da utilização de serviços de *Cloud Computing*.

No projeto *Cloud Service Broker* foi desenvolvido um recomendador e uma camada de abstracção que agrega diversas soluções comerciais de PaaS. O desenho da solução envolveu inicialmente uma análise das APIs das diversas plataformas selecionadas bem como a definição das funcionalidades essenciais que deveriam ser suportadas. Os testes realizados revelaram que a arquitetura não introduziu um *overhead* significativo na maior parte das operações suportadas. Sendo assim, interagir com o *PaaSManager* não coloca mais complexidade nem origina um muito maior tempo de resposta em comparação a interagir diretamente com cada fornecedor.

Em suma, o protótipo desenvolvido no projeto *Cloud Service Broker* é neste momento um dos únicos nesta área de investigação de interoperabilidade e recomendação em ambientes PaaS que possui uma implementação estável e com resultados operacionais. Recentemente, a temática discutida ao longo deste trabalho tem vindo a receber uma grande atenção por parte da comunidade, esperando-se novas iniciativas e projetos que tencionem dar aos utilizadores a oportunidade de controlar várias plataformas de forma unificada.

REFERÊNCIAS

[1] M. Armbrust et al., "A View of Cloud Computing," *Commun. ACM*, vol. 53, pp. 50-58, 2010.

[2] D. Beimborn et al., "Platform as a Service," *Business & Information Systems Engineering*, vol. 3, pp. 381-384, 2011.

[3] S. Ried, "Platform-as-a-service market sizing," 2009.

[4] CloudBees. [Online]. Available: http://www.cloudbees.com/.

[5] CloudFoundry. [Online]. Available: http://www.cloudfoundry.com/.

[6] Heroku. [Online]. Available: https://www.heroku.com/.

[7] Q. Shao et al., "A Performance Guarantee Approach for Cloud Applications Based on Monitoring," em *Proceedings of the 2011 IEEE 35th Annual Computer Software and Applications Conference Workshops*, Washington, DC, USA, 2011.

[8] T. a. E. V. C. a. M. M. a. B. I. Mastelic, "M4Cloud - Generic Application Level Monitoring for Resource-shared Cloud Environments," em *Proceedings of the 2nd International Conference on Cloud Computing and Services Science*, Oporto, Portugal, 2012.

[9] C-CAST. [Online]. Available: http://www.ict-ccast.eu/.

[10] Cloud4SOA. [Online]. Available: http://www.cloud4soa.eu/.

CVS DOS AUTORES

David Cunha, M.Sc. em Engenharia de Redes e Serviços pela Universidade do Minho em 2012. Iniciou a sua actividade profissional na Portugal Telecom Inovação em 2011. Os seus interesses contemplam o open-source, desenvolvimento para a web, bem como as temáticas da *cloud*: Platform-as--a-Service e Software-as-a-Service.

Pedro Neves, Doutorado e Mestre em Engenharia Electrónica, Telecomunicações e Informática pela Universidade de Aveiro em 2012 e 2006, respetivamente. Em 2003 tornou-se bolseiro de investigação do Instituto de Telecomunicações, onde trabalhou em projetos co-financiados pela Comissão Europeia na área de mobilidade e qualidade de serviço. Em Junho de 2006 iniciou atividade na PT Inovação na mesma área de trabalho. Desde 2010 que acumula também atividades de investigação na área de Cloud Computing. Participou em mais de 10 projetos de colaboração internacional, é co-autor de 6 livros internacionais e de mais de 30 artigos publicados em revistas e conferências internacionais.

Pedro Sousa, Doutorado e Mestre em Ciências da Computação pela Universidade do Minho em 2005 e 1997, respetivamente. Em 1996, juntou-se ao Grupo de Comunicações e Redes de Computadores do Departamento de Informática da Universidade do Minho, onde é Professor Assistente e executa suas atividades de investigação no Centro Algoritmi.

05 WebRTC


PAULO CHAINHO


SÉRGIO FREIRE


TELMA MOTA


VASCO AMARAL

PALAVRAS CHAVE
WebRTC, HTML5, VoIP

A tecnologia WebRTC (*Web Real-Time Communications*) promete revolucionar o mercado das Comunicações colocando novos desafios ao negócio das Telecomunicações e gerando novas oportunidades. No entanto para que tal aconteça, é necessário a PT preparar-se e antecipar-se ao mercado com novas ofertas baseadas em novos paradigmas arquiteturais e tecnológicos. Este artigo procura contribuir com algumas recomendações e propostas, começando por caracterizar a tecnologia, os riscos e as oportunidades de negócio, analisando subsequentemente as melhores opções técnicas para a construção de ofertas WebRTC. No final, são apresentados os trabalhos em curso na PT Inovação para validar experimentalmente algumas das opções introduzidas neste artigo e são elaboradas algumas recomendações.

1. INTRODUÇÃO

A tecnologia WebRTC é o próximo grande marco da indústria das comunicações, após pouco mais duma década de Voz sobre IP (*Voice Over IP - VoIP*) no mercado. Existe neste momento um grande burburinho à volta desta tecnologia, gerando grandes expectativas no seu efeito disruptivo. Naturalmente, haverá algum exagero e especulação associado ao tema mas, neste caso, e na opinião dos autores deste artigo, o burburinho justifica-se. É urgente que fornecedores de serviços de comunicação (CSP) como a PT estejam preparados para mitigar os riscos emergentes mas, acima de tudo, se preparem para aproveitar as grandes oportunidades que a tecnologia WebRTC oferece ao mercado das comunicações.

A tecnologia WebRTC permite a comunicação em tempo real entre *browsers*, sem a necessidade de instalar aplicações ou plug-ins adicionais. Consequentemente, qualquer dispositivo que tenha um *browser* será capaz de suportar, de forma nativa, voz e vídeo ou qualquer outro serviço, em tempo real (i.e. conferência, chamadas de voz/vídeo e jogos). Para além disso, as aplicações WebRTC terão a maior parte da sua lógica executada no dispositivo cliente (*browser*), necessitando duma infraestrutura de rede muito simples, especialmente se for comparada com as redes IMS (IP *Multimedia Subsystem*). Por outro lado, o desenvolvimento de novas aplicações de comunicação e colaboração, ricas em funcionalidades e integradas num ambiente *web*, serão muito mais simples de desenvolver e estarão acessíveis a uma comunidade muito grande de programadores *Web*. Isto significa que o custo por utilizador para uma solução 100% baseada em tecnologias *Web* será muito inferior a uma solução baseada nas tecnologias IMS.

Este artigo procura caracterizar a tecnologia WebRTC e o seu estado de disponibilidade nas implementações existentes (Secção 2), bem como analisar o impacto que esta tecnologia pode ter no negócio de um fornecedor de serviços como a Portugal Telecom, tanto em termos de riscos como de oportunidades (Secção 3). São apresentadas várias opções arquiteturais para a oferta de soluções WebRTC (Secção 4) e as experimentações em curso nos laboratórios da PT Inovação para avaliar algumas dessas opções (secção 5). Finalmente, a secção 6 sumariza as principais conclusões e recomendações.

2. A TECNOLOGIA

2.1 INTRODUÇÃO

WebRTC é uma tecnologia emergente que permite comunicações em tempo real entre navegadores *web*. Para muitos é vista como uma *framework JavaScript* que se conjuga com as APIs HTML5, nomeadamente com elementos de áudio e vídeo, permitindo aos programadores *web* criar facilmente aplicações com comunicação multimédia ponto-a-ponto, sem requerer plugins proprietários no navegador local.

A API *Javascript* fornece um conjunto de funcionalidades base que permitem o acesso a dispositivos locais multimédia e o estabelecimento de *streams* de áudio, vídeo e dados entre navegadores. Pode-se dividir as funcionalidades de WebRTC em três grandes APIs [1]:

GetUserMedia

Permite:

- Acesso aos recursos multimédia do utilizador, nomeadamente câmara e microfone.

>>>

- Obter um *stream* que pode ser usado como fonte de um determinado elemento, e.g. vídeo, ou ser transmitido através duma ligação ponto-a-ponto (i.e. *PeerConnection*).

PeerConnection

Permite:

- Estabelecer sessões de comunicação entre navegadores.
- NAT (*Network Address Translation*) usando ICE (*Internet Connectivity Establishment*) com STUN/ TURN (*Session Traversal Utilities/ Traversal Using Relays around NAT*).
- Negociar *codecs* através de SDP (*Session Description Protocol*), com um conjunto mandatório mínimo de codecs por forma a garantir interoperabilidade - G.711, G.722, iLBC and iSAC para áudio, e VP8 para vídeo.
- Adaptação à largura de banda da ligação.
- Cancelamento de eco e ajuste de áudio.

DataChannel

Permite:

- Troca de dados arbitrários (e.g. binários) ponto-a-ponto.
- Comunicações seguras e inseguras baseadas no protocolo SCTP (*Stream Control Transmission Protocol*) e no DTLS (*Datagram Transport Layer Security*) para garantir segurança.
- Possibilidade de garantia de entrega ordenada de pacotes.
- Múltiplos canais em simultâneo.

Uma funcionalidade não endereçada pelo WebRTC é a de sinalização e respetivo protocolo para estabelecimento e negociação da comunicação entre os *peers*; cabe ao programador desenvolver mecanismos que permitam aos navegadores descobrirem-se e trocar informações entre si – tipicamente isto pode ser feito usando um servidor intermediário e tecnologia HTML5 *WebSockets*, ou até mesmo um *DataChannel*.

Muita da tecnologia por detrás do WebRTC foi desenvolvida pela *Global IP Solutions* (GIPS), que no passado havia fornecido parte da sua tecnologia para o Skype, antes do Skype ter desenvolvido as suas próprias soluções. A Google entretanto adquiriu a GIPS em Maio de 2010 e tornou público o código fonte da *framework* intitulada WebRTC em Junho de 2011. A Google e outras organizações como a Mozilla, Opera, Ericsson e Cisco continuam a esforçar-se pela normalização do WebRTC no W3C (*World Wide Web Consortium*) e no IETF (*Internet Engineering Task Force*). Inicialmente, a recomendação estava prevista para o 1º trimestre de 2013 mas a alternativa elaborada pela Microsoft, i.e. CU-RTC-Web (*Customizable, Ubiquitous Real-Time Communication over the Web*, 2013), em Junho de 2012 [2], atrasou o processo e gerou alguma discussão, devido a diferenças substanciais de abordagem.

O WebRTC, a ser adotado, irá ter um grande impacto na forma como as pessoas decidem comunicar, especialmente nas empresas. É muito provável que vejamos os sítios (*sites*) das companhias usar o WebRTC como o método preferencial de contato, por forma a chegar a mais pessoas e a reduzir custos. Quando combinado com a autenticação do utilizador e o seu contexto no sítio *web*, o operador do *call center* terá acesso imediato a detalhes relevantes, reduzindo dessa forma o tempo de atendimento e melhorando a experiência de utilização. Os sítios das redes sociais, como o Facebook, integrarão possivelmente o WebRTC nas suas funcionalidades de conversação *online*, como forma de fornecer instantaneamente capacidade de interação áudio e vídeo. Se aliarmos a isto algumas das APIs HTML5 anteriormente descritas e outras relacionadas, teremos aplicações *web* extremamente interessantes e inovadoras que podem incluir funcionalidades como o controlo por gestos, verificação de utilizador, interação por voz, etc.

2.2 ARQUITETURA DE REFERÊNCIA

A arquitetura de referência atual pode ser visualizada na figura seguinte:

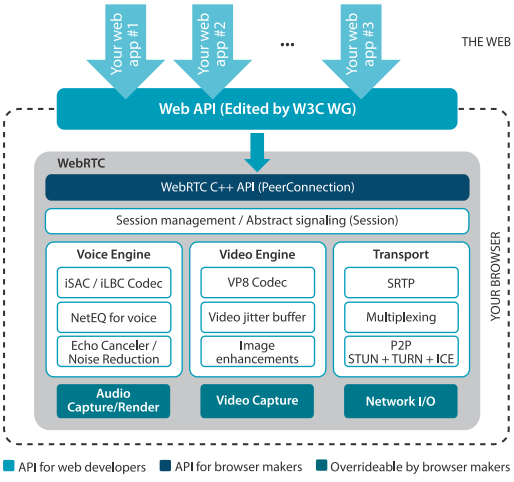


Figura 1. Arquitetura de referência [3]

2.3 APIS

Estão a ser desenvolvidas dois tipos diferentes de APIs.

A primeira e mais importante do ponto de vista da comunidade de programadores é a chamada *API user-side*, também chamada de "*WebRTC - Real-time Communication Between Browsers*". É esta que está disponível para os programadores *web* desenvolverem páginas ou aplicações *web* com capacidades RTC.

A API "*WebRTC Native C++*", ao contrário da primeira, é indicada para fornecedores de *browsers*, que queiram implementar uma camada de abstração de acesso às funcionalidades da API *Javascript*. A sua especificação baseia-se inteiramente na especificação da API user-side, estando por isso estreitamente interligadas.

2.4 PILHA PROTOCOLAR

Na figura seguinte é possível comparar os protocolos tipicamente usados nas aplicações *web* (lado esquerdo) com os protocolos envolvidos no WebRTC (lado direito).

A pilha protocolar WebRTC assenta no protocolo UDP. Acima deste podemos ter o STUN ou o TURN para lidar com questões de NAT e *firewall*, servindo o ICE para coordenar os primeiros. A segurança é assegurada pelo DTLS (*Datagram Transport Layer Security*) ou então pelo SRTP (*Secure Real-time Transport Protocol*), no caso dos pacotes media. Como meio de comunicação existem duas entidades, a *PeerConnection* e o *DataChannel*, sendo o primeiro focado na comunicação de *media* usando o SRTP, enquanto o *DataChannel* usa o SCTP para fornecer as funcionalidades relacionadas com multicanal, entrega ordenada e assegurada.

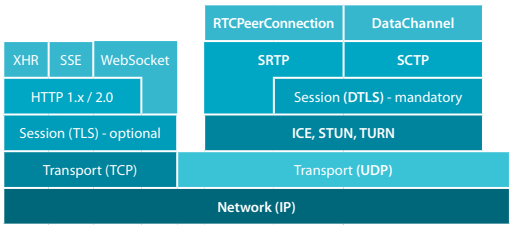


Figura 2. Pilha Protocolar WebRTC (lado direito) vs. aplicações *web* típicas (lado esquerdo)

2.5 COMUNICAÇÃO *MEDIA*

A figura seguinte representa as entidades envolvidas numa hipotética comunicação multimédia entre dois *browsers* "A" e "B". Um *Media Stream* pode ser composto por um ou mais *tracks* provenientes de um ou mais dispositivos multimédia locais. Os vários *Media Streams* são encapsulados numa sessão RTP (SRTP na realidade). Tudo isto é possível usando uma *PeerConnection* estabelecida entre "A" e "B".

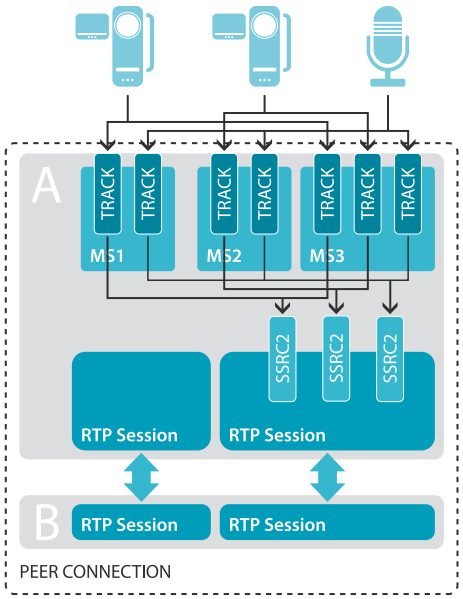


Figura 3. Entidades WebRTC envolvidas numa sessão multimédia

MSx: *Media Stream* (conjunto de pacotes gerados por uma fonte de media – microfone, câmara, ...)

SSRC: *Sender Source* (identificador 32-bits atribuído por sessão RTP)

2.6 ESTADO ATUAL DA TECNOLOGIA

A tecnologia WebRTC começou pelo suporte da API *getUserMedia*, seguida da *PeerConnection*. A API *DataChannel*, mais recente, encontra-se ainda em desenvolvimento.

Já foi demonstrada interoperabilidade entre os *browsers* Chrome e o Firefox, nomeadamente nas APIs *getUserMedia* e *PeerConnection*. No entanto, o suporte completo do WebRTC ainda não existe em nenhum navegador, ainda que o Chrome e o Firefox implementem uma grande parte da especificação atual, com limitações na componente *DataChannel*. A Microsoft insiste na alternativa CU-RTC-Web, o que está a atrasar a adoção da tecnologia no Internet Explorer nas várias plataformas Windows.

Entretanto surgiu mais uma alternativa – ORTC API (Object RTC API for WebRTC) [4], em Agosto de 2013. Esta abordagem pretende endereçar alguns dos problemas da API WebRTC, nomeadamente o modelo SDP *Offer/Answer* que se demonstra desadequado numa série de cenários, além de estar de alguma forma virado para o mundo SIP tradicional. A ORTC API é, tal como o *draft* CU-RTC-Web, uma API de mais baixo nível, sendo em teoria possível, com base nas suas primitivas, construir em *Javascript* uma API compatível com o WebRTC atual.

>>>

Face às diferentes propostas e à discussão levantada pela comunidade, a Google acabou por concordar com a necessidade da existência duma API de mais baixo nível, a ser implementada numa versão posterior do WebRTC [5]. Entretanto o foco, até finais de 2013, passa por definir o WebRTC 1.0.

Nos dispositivos móveis, o suporte é mais limitado e tem acontecido de forma mais moderada. Contudo é expectável que agora que o suporte está estabilizado nos *browsers desktop*, as funcionalidades venham a ser ativadas nos *browsers* móveis. A grande questão está nos dispositivos *iOS* da Apple, que devido a restrições impostas pela marca impedem o uso de um motor webkit diferente do fornecido pela Apple, limitando quer o Chrome (que é basea-do no *webkit* mas mais recente e completo) quer outros *browsers*. A Ericsson contudo demonstrou, num protótipo intitulado *iLeif*, que é possível estender o *webkit* do *iOS* por forma a suportar novas APIs como o WebRTC.

2.7 SUPORTE NOS *BROWSERS*

A tabela seguinte dá um panorama do estado da arte à data do 2º trimestre de 2013 [6], indicando o suporte das três interfaces/funcionalidades macro.

	getUserMedia	*PeerConnection*	*DataChannel*	Observações
Chrome v28	Sim	Sim	Parcial	*getUserMedia*: desde a v.21 (estável) ativa sem a *flag* em "chrome://flags" apenas *DataChannel* do tipo "*unreliable*" • *DataChannel* experimental e interoperável com Firefox desde v26
Chrome mobile Beta (v.29)	Sim	Sim	Parcial	http://blog.chromium.org/2013/07/chrome-29-beta-web-audio-and-webrtc-in.html
Opera Desktop (v.12)	Sim	Não	Não	Não usa prefixo do *browser*.
Opera Desktop (v.15)	Não	Não	Não	Ainda que baseado em código do Chrome 28, não tem o WebRTC ativo.
Opera Mobile (v.12, Android)	Sim	Não	Não	Não usa prefixo do *browser*.
Firefox Desktop (v.22)	Sim	Sim	Sim	Não suporta TURN.
Firefox Mobile	Não	Não	Não	Possivelmente na v24. (https://hacks.mozilla.org/2013/06/webrtc-comes-to-firefox/)
Internet Explorer 9, 10	Não	Não	Não	Sem suporte previsto. A MS está a promover o CU-RTC-Web.
Safari (OSX & iOS5)	Não	Não	Não	Sem suporte previsto.

3. IMPACTO NO NEGÓCIO DAS TELECOMUNICAÇÕES

3.1 O RISCO *OVER-THE-TOP*

A tecnologia WebRTC tem como grande impacto no negócio das telecomunicações a aceleração da tendência da perca de negócio para os grandes intervenientes *Over-The-Top* (OTT) como é o caso da Google, Facebook, ou WhatApps ou outros que entretanto surjam com novas aplicações e modelos de negócio disruptivos.

3.1.1 Exemplos de Serviços WebRTC OTT

Apesar de ser uma tecnologia muito recente, pouco amadurecida e cujas normas ainda não estão finalizadas, já existe um número considerável de serviços baseados em WebRTC que são alimentados pela criatividade duma enorme comunidade de programadores *web*. Grande parte destes serviços é de empresas *startup* muito focadas no uso da tecnologia WebRTC. Até ao momento, a adoção desta tecnologia por parte dos OTT estabelecidos é reduzida.

Os serviços focados numa comunicação básica áudio e vídeo *a la Skype* ou do tipo colaborativo *a la Webex/Go To Meeting* são os primeiros tipos de serviço onde o uso da tecnologia WebRTC é mais óbvio e mais simples. Existem inúmeros exemplos destes serviços como é o caso do Talky [7] e do UberConference [8] com funcionalidades de audio, vídeo e *chat* para grupos de utilizadores, bem como funcionalidades avançadas de *Screen Sharing*, Gravação e *Floor-control*.

Alguns destes serviços, como o Oscar [9] ou o Tawk [10], destacam-se pela sua simplicidade e experiência de utilização imersiva, apostando num desenho

muito cuidado da Interface de Utilizador. Por exemplo, no serviço Oscar as conversações de áudio e vídeo são apresentadas e manipuladas como se estivessem a ocorrer num local virtual escolhido pelos utilizadores (e.g. num parque verde) aliado a um grafismo de grande qualidade.

Outros serviços, como o Bistri (ref. [11] ou o Twelephone [12], procuram promover a sua utilização através duma integração muito próxima com os serviços de redes sociais existentes (e.g. Twitter, Facebook, Google) usando as identidades dos utilizadores desses serviços e adicionando alguns efeitos especiais sobre as imagens de vídeo dos utilizadores para tornar mais divertida a sua experiência.

Outra utilização imediata da tecnologia WebRTC é a possibilidade de a disponibilizar em portais *web*, em particular Portais de Comércio Eletrónico, capacidades do tipo *click to talk* para assistência aos visitantes destes portais. Zingaya [13] e Teledini [14], são apenas alguns exemplos de serviços deste tipo que já usam duma forma comercial a tecnologia WebRTC.

A interação e colaboração no âmbito dos serviços de Redes Sociais também beneficiam muito do uso da tecnologia WebRTC como é o caso do serviço Solaborate [15] ou do serviço Maxiamigos [16], sendo este último um serviço de encontros Brasileiro.

Os exemplos até agora referidos endereçam mercados e negócios que já existiam antes de aparecer a tecnologia WebRTC e que beneficiam desta tecnologia quer pelos custos de desenvolvimento e operação quer pela melhoria da experiência do utilizador.

No entanto, começam a surgir novas oportunidades de negócio alavancadas pelo uso da tecnologia WebRTC, com destaque para o mercado dos "Especialistas". Neste novo mercado, é promovido o encontro entre Especialistas em certas áreas e os utilizadores que precisem da sua ajuda e que, eventualmente, estão dispostos a pagar para ter uma sessão de Vídeo exclusiva com algum destes especialistas. Sessões de terapia ou de exercício físico, aulas ou explicações sobre determinada matéria, consultadoria ou assistência nalguma atividade profissional são apenas alguns exemplos. Estes serviços também são uma excelente oportunidade para os especialistas obterem receitas adicionais com poucos ou nenhuns custos, num mercado potencialmente global. Esta característica torna estes serviços especialmente atraentes para especialistas residentes em mercados com grande desemprego e com uma taxa de emigração elevada, como é infelizmente o caso Português. Liveninja [17] e Popexpert [18] são boas referências de serviços de Especialistas.

Existem outros tipos de serviços mais técnicos que tiram partido das capacidades do *DataChannel* do WebRTC como é o caso do serviço PeerCDN [19]. Neste serviço, o *DataChannel* é usado no estabelecimento de redes de fornecimento de conteúdos CDN (*Content Delivery Network*), extremo a extremo; os conteúdos de um portal podem circular diretamente entre os *browsers* dos seus visitantes sem consumir recursos dos servidores do portal.

Finalmente, começam a surgir os primeiros serviços móveis baseados em tecnologia WebRTC com destaque para o Mobile Vonage [20] já com alguns milhões de utilizadores. Este serviço disponibiliza gratuitamente chamadas de áudio, serviço de mensagens de texto e partilha de fotografias para *Smartphones iPhone* e *Android*.

3.2. OPORTUNIDADES

Os exemplos de serviços apresentados na secção anterior dão uma ideia do potencial de utilização da tecnologia WebRTC que muito rapidamente proporcionará criar no mercado novos serviços capazes de retirar negócio aos CSPs (*Communication Service Provider*) como a PT. No entanto, pode criar também muitas oportunidades que, no limite, poderão compensar os riscos referidos anteriormente.

A primeira grande oportunidade surge na possibilidade dos CSPs fornecerem interoperabilidade de voz e vídeo com os terminais existentes, com destaque para os dispositivos móveis. Este negócio poderá ser bastante lucrativo nos próximos anos mas, com a expansão da tecnologia WebRTC, terá tendência a diminuir. A operadora norte-americana AT&T é uma boa referência no modo como já está a explorar esta oportunidade através da oferta duma biblioteca *javascript* att.js [21] que facilita a inclusão nos portais *web* das funcionalidades de *click to dial* para qualquer número de telefone da rede AT&T.

A rentabilização da infraestrutura existente, como o IMS, para criar novas ofertas de produtos ou expandir ofertas existentes para terminais WebRTC, é outra possibilidade a explorar a curto prazo. Mas também aqui a tendência será estas soluções perderem rapidamente a sua competitividade quando comparadas com soluções desenvolvidas de raiz por novos intervenientes *startups* ou OTTs já bem estabelecidos, que sejam capazes de aproveitar todo o potencial das tecnologias *web* e da sua comunidade de programadores. Nos últimos meses,

>>>

o mercado tem sido inundado com anúncios de grandes fornecedores de soluções e.g. *Huawei Web RCS Gateway* [22], *Ericsson Web Communication Gateway* [23], *GenBand WebRTC Gateway* [24], empenhados em explorar esta oportunidade, assim como de grandes Operadores atarefados nos seus laboratórios a testar estas soluções.

Outra grande oportunidade reside na criação de ofertas de produtos em segmentos verticais como a saúde e a educação, baseados em soluções 100% *web*. A solução Medigraf da PT Inovação é um bom exemplo de como esta abordagem pode ser usada com sucesso.

Finalmente, existe a grande oportunidade da indústria em geral repensar a sua infraestrutura e usar uma nova abordagem baseada, em grande parte, em tecnologias *web*. O objetivo é criar um ecossistema de serviços capaz de suportar um comportamento mais ágil e inovador na disponibilização no mercado de novas ofertas competitivas e à medida das suas necessidades. A aquisição da TokBox [25] pela Telefonica é uma decisão estratégica que aponta nessa direcção. A TokBox é uma *startup* tecnológica líder no uso da tecnologia WebRTC para a construção de novas soluções como é o caso do serviço Oscar.

4. ARQUITECTURAS PARA O FORNECIMENTO DE SERVIÇOS WEBRTC

No estudo Eurescom P2252 [26], foram analisadas quatro opções diferentes para o fornecimento de serviços WebRTC, usando:

1. Servidores Aplicacionais SIP onde se aproveita a infraestrutura IMS existente.

2. Servidores Aplicacionais *Web* numa abordagem baseada em tecnologias *web*.

3. Plataformas VoIP não compatíveis com a arquitectura IMS, como é o caso de plataformas *Open Source Asterisk, Freswitch* e *OpenSER*.

4. APIs num ambiente orientado-para-serviços (*service oriented*) onde a natureza disruptiva do WebRTC é usada como uma oportunidade para a indústria das comunicações dar um passo em frente fazendo evoluir o seu *core*, tornando-se mais ágil na sua base através da aplicação de paradigmas arquiteturais SOA na camada de controlo das redes de Telecomunicações.

Vamos analisar com mais cuidado as opções 1 e 2 que serão as opções mais viáveis a curto prazo.

4.1 OFERTAS BASEADAS EM IMS

As ofertas WebRTC baseadas em IMS assentam em infraestruturas IMS existentes sendo particularmente apropriadas para estender a terminais WebRTC o fornecimento de serviços tradicionais de voz, como serviços Classe 5 e IP-Centrex e serviços multimédia mais avançados como RCS/Joyn e o MMTEL. O elemento central numa oferta de serviços baseada no IMS (figura 4) é o *Media Access Gateway* que será responsável por traduzir a sinalização usada nos terminais WebRTC (tipicamente mensagens SIP transportadas em *Web Sockets*) na sinalização IMS SIP. Adicionalmente, o *Media Access Gateway* deverá também lidar com os mecanismos de encriptação dos *streams* de *media* WebRTC e eventualmente com a sua transcodificação, caso sejam usados *codecs* diferentes, e.g. VP8 vs H.248. Para um fornecedor de serviços que tenha um catálogo rico em serviços IMS, deverá bastar uma atualização dos *Media Access Gateway* (*Session Border Controller*) já existentes na rede para os tornar acessíveis a partir de terminais com capacidades WebRTC.

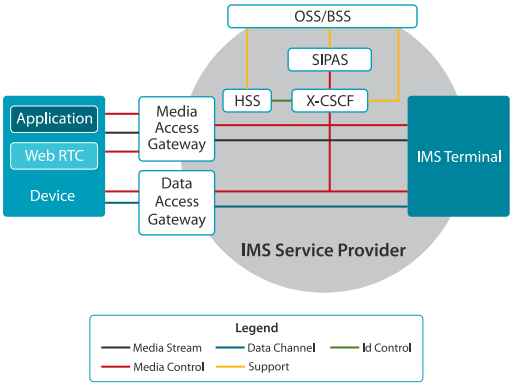


Figura 4. Arquitetura para uma oferta WebRTC baseada em soluções IMS

4.2 OFERTAS BASEADAS EM SERVIDORES *WEB*

Nesta opção, o desenvolvimento e fornecimento de serviços é inteiramente baseado em tecnologias *web* tirando proveito da enorme quantidade de Servidores Aplicacionais (ASs) *web* instalados e da grande comunidade de programadores *web*. Neste caso (ver Figura 5) o servidor aplicacional *web* é o principal elemento de rede fornecendo a sinalização e outras funcionalidades avançadas como serviços de PBX virtuais. Adicionalmente, pode existir um *Media Application Server* que disponibiliza funcionalidades de processamento de media, e.g. mistura de fluxos de *media* para vídeo-conferência. Nesta opção, o protocolo de sinalização não está normalizado, isto é, a interoperabilidade só é garantida se os clientes WebRTC estiverem ligados no mesmo *web AS*. Serão

>>>

necessárias *gateways* para garantir interoperabilidade entre clientes WebRTC e outros tipos de clientes e.g., IMS, PSTN/PLMN. Do ponto de vista de segurança, as entidades envolvidas numa sessão podem ser autenticadas por um *Identity Provider* externo baseado em tecnologias *web* (e.g. OpenId ou OAuth). Finalmente, para ter funcionalidades de QoS e de Faturação, o *Web AS* pode interagir com sistemas OSS/BSS usando por exemplo simples APIs REST.

Para tirar partido da total convergência entre serviços mais orientados aos dados (*web*) e serviços mais orientados à comunicação, serviços como PBX ou Comunicações Unificadas deverão ser desenhados e desenvolvidos de raiz de modo a promover uma nova classe de serviços.

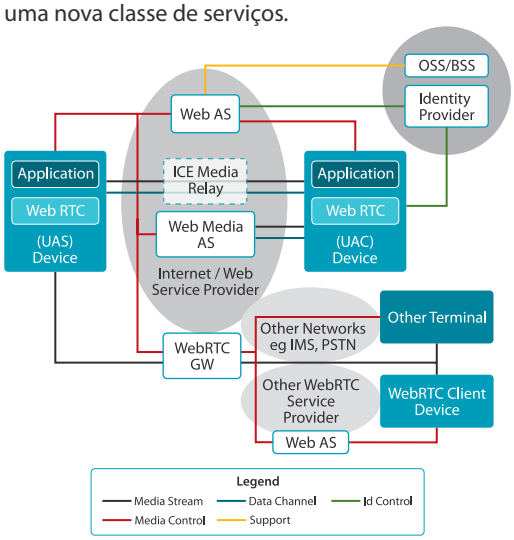


Figura 5. Arquitetura duma oferta WebRTC baseada em Servidores *Web*

5. PROTÓTIPO

A PT Inovação tem em curso um conjunto de experimentações para explorar as duas principais opções de oferta de serviços anteriormente descritas.

No âmbito das ofertas baseadas em IMS, a PT Inovação tem no seu laboratório uma solução da *ASC ACME Packet* [27] que está a testar com um conjunto de clientes WebRTC, com destaque para os clientes *open source SIPML5* [28] e SIPPO [29]. Em paralelo está a decorrer o desenvolvimento dum cliente WebRTC para o serviço SEC (Serviços Empresariais Convergentes).

Numa linha de trabalhos de Inovação Exploratória e dentro das ofertas de serviços baseados em servidores aplicacionais *web*, a PT Inovação desenvolveu uma prova de conceito que explora todas as capacidades da tecnologia WebRTC, usando novos paradigmas de comunicação orientada "a eventos" que abrange funcionalidades de cliente e de servidor. Do trabalho de alguns meses, usando novas abordagens baseadas em tecnologias *web*, resultou uma solução rica em funcionalidades incluindo (ver Figura 6):

1. gestão de lista de contactos enriquecida com informação de presença;
2. estabelecimento de sessões a partir da lista de contactos, por descoberta no diretório de utilizadores aprovisionados ou através de um link de acesso que pode ser usado por utilizadores não aprovisionados e enviado *out of band*, e.g. por *email* ou *chat*;
3. sessões multi-utilizador de áudio e vídeo enriquecidas com funcionalidades de *chat*, partilha de ficheiros e de ecrã;
4. histórico de sessões.

De destacar que as funcionalidades de *Chat*, Partilha de Ficheiros e de Ecrã são suportadas nos canais de dados (*DataChannel*) do WebRTC que permitem a transferência de dados extremo a extremo sem consumir recursos de servidores do lado da rede.

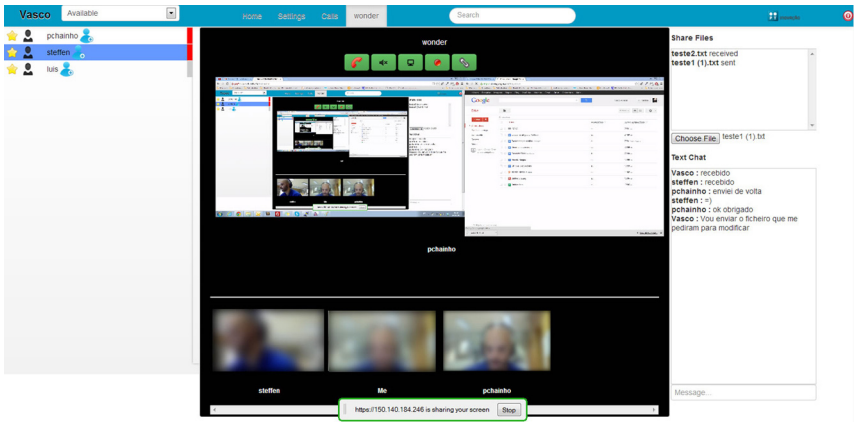


Figura 6. Prova de Conceito de Aplicação WebRTC da PT Inovação

6. RECOMENDAÇÕES E CONCLUSÕES

Este artigo procurou caracterizar a tecnologia WebRTC e o seu estado de disponibilidade nas implementações existentes, assim como demonstrar que já é possível o desenvolvimento de soluções experimentais e o lançamento de pilotos para avaliar no terreno o seu potencial.

Foi também analisado o impacto que esta tecnologia pode ter no negócio de um fornecedor de serviços como a Portugal Telecom, em termos de riscos e oportunidades. Neste âmbito, foram apresentados alguns exemplos dos inúmeros serviços já existentes no mercado suportados na tecnologia *web*, bem como algumas iniciativas de CSPs estabelecidos. Foram apresentadas várias maneiras de mitigar o grande risco de aceleramento de perca de receitas para fornecedores OTT mas, acima de tudo, procurou salientar-se a grande oportunidade que o WebRTC pode ser para a indústria das Telecomunicações: desenhar uma nova infraestrutura capaz de sustentar um comportamento mais ágil e inovador no mercado.

Foram apresentadas várias opções arquiteturais para a oferta de soluções WebRTC. Para CSPs que já tenham uma infraestrutura IMS rica em serviços como MMTEL, VoLTE, Centrex e RCS, a opção natural, no curto prazo, é reutilizar estes ativos e estender a sua oferta para terminais com capacidades WebRTC. No entanto, tendo em conta que a tecnologia WebRTC potencia gerar muito tráfego, esta opção poderá implicar a necessidade de aumentar a capacidade da infra-estrutura IMS. Isto significa que os CSPs terão de construir casos de negócio fortes e sustentáveis para justificar os investimentos adicionais na infra-estrutura IMS. Nesta abordagem, o maior desafio é manter uma oferta competitiva com a competição OTT, que usa infraestruturas mais simples e económicas onde o custo por cliente WebRTC é inferior ao custo por cada cliente IMS. Para enfrentar este desafio os CSP podem seguir uma abordagem híbrida onde complementam a sua oferta baseada em IMS com uma oferta baseada em servidores aplicacionais *web*, capaz de competir directamente com a concorrência OTT. A médio e longo prazo os fornecedores devem evoluir a sua oferta de acordo com os princípios da orientação-ao-serviço e governada pelos melhores processos e políticas de negócio praticadas no mercado. Deste modo será possível ter capacidade para introduzir serviços mais sofisticados e convergentes, verdadeiramente inovadores que ofereçam aos consumidores uma experiência holística e imersiva.

Seguindo esta estratégia, os resultados iniciais das provas de conceito da PT Inovação são promissores. Actualmente, esta solução está a ser validada e melhorada em parceria com a Deutsche Telekom no âmbito do projeto Europeu WONDER, parcialmente financiado pela Comissão Europeia. Em particular, este projecto ambiciona avaliar as duas opções de construção de serviços WebRTC apresentadas neste artigo (baseada em IMS vs baseada em servidores *web* bem como experimentar soluções de interoperabilidade entre diferentes domínios WebRTC baseados em APIs. Os resultados desta experimentação estão previstos para o início de 2014.

REFERÊNCIAS

[1] WebRTC 1.0: Real-time Communication Between Browsers. (August de 2013). Obtido em September de 2013, de W3C: http://dev.w3.org/2011/webrtc/editor/webrtc.html
[2] CU-RTC-WEB: Customizable, Ubiquitous Real-Time Communication over the Web. (April de 2013). Obtido em September de 2013, de W3C: http://lists.w3.org/Archives/Public/public-webrtc/2012Oct/att-0076/realtime-media.html
[3] WebRTC Architecture. (s.d.). Obtido em September de 2013, de http://www.webrtc.org/reference/architecture
[4] Object RTC (ORTC) API for WebRTC. (s.d.). Obtido em September de 2013, de https://github.com/openpeer/ortc/blob/master/draft-w3c-ortc-api-00.md
[5] Google's position on the WebRTC API. (s.d.). Obtido em September de 2013, de http://lists.w3.org/Archives/Public/public-webrtc/2013Jul/0484.html
[6] WebRTC support summary. (s.d.). Obtido em September de 2013, de http://www.html5rocks.com/en/tutorials/webrtc/basics/#toc-support
[7] Talky. (s.d.). Obtido em September de 2013, de https://talky.io/
[8] uberconference. (s.d.). Obtido em September de 2013, de http://www.uberconference.com/
[9] OSCAR. (s.d.). Obtido em September de 2013, de http://oscar.tokbox.com
[10] tawk. (s.d.). Obtido em September de 2013, de https://tawk.com/
[11] Bistri. (s.d.). Obtido em September de 2013, de https://bistri.com/
[12] Twelephone. (s.d.). Obtido em September de 2013, de http://twelephone.com/
[13] Zingaya. (s.d.). Obtido em September de 2013, de http://zingaya.com/
[14] Teledini. (s.d.). Obtido em September de 2013, de http://www.teledini.com/
[15] Solaborate. (s.d.). Obtido em September de 2013, de https://www.solaborate.com
[16] MaxiAmigos. (s.d.). Obtido em September de 2013, de http://maxiamigos.com/
[17] Liveninja. (s.d.). Obtido em September de 2013, de https://www.liveninja.com/
[18] PopExpert. (s.d.). Obtido em September de 2013, de https://www.popexpert.com
[19] PeerCDN. (s.d.). Obtido em September de 2013, de https://peercdn.com/
[20] Vonage Mobile. (s.d.). Obtido em September de 2013, de http://www.vonagemobile.com/
[21] att.js. (s.d.). Obtido em September de 2013, de https://js.att.io
[22] Huawei Web RCS Gateway. (s.d.). Obtido em September de 2013, de http://pr.huawei.com/en/news/hw-259848-webrtc.htm
[23] Ericsson Web Communication Gateway. (s.d.). Obtido em September de 2013, de http://www.ericsson.com/us/ourportfolio/products/web-communication-gateway
[24] GenBand WebRTC Gateway. (s.d.). Obtido em September de 2013, de http://www.genband.com/products/experius/webrtc-gateway
[25] TokBox. (s.d.). Obtido em September de 2013, de http://tokbox.com/
[26] Eurescom. (2012). P2252 - Telco strategic positioning options regarding WebRTC. Eurescom.
[27] Acme Packet ASC. (s.d.). Obtido em September de 2013, de http://www.acmepacket.com/products-services/service-provider-products/net-net-application-session-controller
[28] Doubango Telecom. (s.d.). SIPML5. Obtido em September de 2013, de http://sipml5.org/
[29] Quobis. (s.d.). SIPPO. Obtido em September de 2013, de http://www.quobis.com/index.php?option=com_content&task=view&id=265&Itemid=196

CVS DOS AUTORES

Paulo Chainho licenciado em Engenharia Electrotécnica e de Computadores pelo Instituto Superior Técnico da Universidade Técnica de Lisboa e Mestre em Telecomunicações pela mesma universidade. Trabalha na PT Inovação desde 2001 na área das Aplicações e Plataformas de Serviços Convergentes para Redes de Nova Geração, incluindo soluções de Servidor de Aplicações SIP e SDP/SDF. Nestas actividades tem desempenhado funções de Gestão de Projecto e de Concepção das soluções. Atualmente participa em atividades de consultoria para desenvolvimento e gestão de produtos SDF e Atividades de Gestão Estratégica. Entusiasta do "Open Source". Grande experiência em projetos Internacionais de Investigação e Desenvolvimento incluindo Eurescom, ETSI SPAN e EU IST.

Sérgio Freire, licenciado e mestrado em Engenharia Eletrónica e Telecomunicações, com pós-graduações em Sistemas de Informação e Redes de Comunicação. Em 2001 iniciou a sua atividade na PT Inovação, tendo até 2011 a área de Messaging (S|MMS, instant e unified messaging) sido o seu enfoque, desde conceção, desenvolvimento e coordenação. Especialista em integração com os mais diversos sistemas e tecnologias. Desde 2012 desempenha funções de consultoria tecnológica. Múltiplos interesses, desde novas tecnologias em geral, HTML5, linguagens de programação, Linux, cloud computing, high-performance, frameworks, networking, protocolos, open-source. Adepto fervoroso da criatividade e da tecnologia. Publicações diversas, incluindo na Usenix. Premiado múltiplas vezes no evento Codebits.

Telma Mota, concluiu a Licenciatura e Mestrado em Engenharia Eletrotécnica e de Computadores na Universidade do Porto. Ingressou na empresa TLP SA, onde realizou trabalho de planeamento e dimensionamento de redes de comutação digital, Redes Inteligentes e teletráfego. Desde 1994 que integra a PT Inovação e tem estado ligada às áreas de gestão e arquiteturas de Redes e Serviços; IN, evolução da IN, TINA, Parlay, IMS, TISPAN e MBMS, assim como às normas 3GPP que se dedicam a definir aspetos de estabelecimento de Sessões Multimédia, QoS, Mobilidade e Multicast. Recentemente tem-se dedicado às arquiteturas de serviços; OMA, SOA, Web 2.0. Participou em diversos projetos Europeus (Eurescom e IST), liderou o C-CAST e na PTIN é responsável pela divisão "Plataformas e Redes Multiserviço".

Vasco Amaral licenciado em Engenharia Informática pela Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Viseu. Aluno do Mestrado em Engenharia Informática na Universidade do Minho, encontrando-se a realizar a dissertação do mestrado na PT Inovação, na qual é estagiário desde Fevereiro de 2013, na área WebRTC.

06 Automatização da Verificação de Políticas de Segurança e Deteção de Vulnerabilidades

MÁRIO MOREIRA RUI BERNARDINO

PALAVRAS CHAVE
Configuração de sistemas, conformidade, política de segurança, perfis, *standards*, *baselines*

O SCAP (*Security Content Automation Protocol*) é um conjunto de protocolos definidos pelo NIST (*National Institute of Standards and Technology*) que pretende criar um formato aberto e expansível para a difusão de conteúdos de segurança. Com estes documentos é possível a criação de *benchmarks*, enumeração de vulnerabilidades, deteção de modificações, avaliação de atualizações, criação de guias, remediação automática, inventariação de recursos, etc. Tanto os inputs como os outputs produzidos são independentes de fabricantes, plataformas e ferramentas, permitindo que políticas idênticas sejam aplicáveis em cenários distintos e com resultados comparáveis.

Este artigo é uma introdução ao SCAP e apresenta um conjunto de ferramentas e políticas criadas pela PT Inovação, alinhadas com a Política de Segurança de Informação do Grupo PT e outros clientes, com vista à avaliação automática de conformidade para soluções assentes em sistemas Red Hat Linux.

1. O QUE É O SCAP

A segurança de sistemas e aplicações é tipicamente um processo complexo, moroso e lento que envolve várias partes e que exige um acompanhamento contínuo e dispendioso.

O ciclo normal da segurança de informação começa nas chefias de topo das instituições: são estas que traçam em linhas gerais os objetivos a atingir no que diz respeito à segurança, possivelmente alinhando com normas internacionalmente reconhecidas (ex. ISO 27001 na segurança de informação ou duma forma mais abrangente à Sarbanes-Oxley), onde são definidos níveis de confidencialidade, qual o risco admissível, quais os agentes responsáveis pelas várias fases, qual a periodicidade de revisões, etc. Tipicamente este documento tem um ciclo de revisão alargado.

Cabe depois às áreas responsáveis pela segurança da informação propagar estas linhas gerais e normas internacionais no que normalmente se designa por Política de Segurança de Informação; nesse documento as linhas diretoras são concretizadas em regras explícitas como por exemplo "*todas as passwords devem ter no mínimo 10 carateres*" ou "toda a informação transmitida pela rede deverá ser cifrada". Como estas regras tomam muitas vezes em conta as condicionantes técnicas do momento em que foram criadas, o seu ciclo de revisão é bastante mais pequeno; pode-se pegar no exemplo das *passwords*, cujos requisitos mínimos têm sido sucessivamente revistos de forma a contrariar o aumento da capacidade de computação e outras evoluções nos métodos criptográficos.

Até agora, referiram-se apenas políticas e regras: documentos, prosa descritiva de normas e procedimentos que devem ser passados aos utilizadores e gestores de plataformas. Estes documentos são essenciais mas não garantem a aplicação prática dos objetivos iniciais da empresa. Se no caso dos utilizadores, o passo seguinte será a divulgação das regras, no caso dos sistemas e aplicações o caso é bastante mais complexo. Alguns problemas surgem nesta fase:

- como validar de forma expedita que um dado sistema está conforme a política de segurança da empresa? Ou do cliente a que se destina?
- como assegurar que as modificações decorrentes da atualização normal dum sistema não vão contra as mesmas políticas? E se o sistema for comprometido por terceiros?
- como uniformizar os critérios de aplicação das regras dos administradores de sistema com os dos auditores?
- como transmitir aos fornecedores externos quais os requisitos aceitáveis no que diz respeito à parametrização de segurança dos sistemas e aplicações?

Foram estas e outras questões que o NIST tentou abordar com a introdução dos protocolos SCAP (*Security Content Automation Protocol*), um protocolo aberto, independente da plataforma, definido em grande parte por contribuições da comunidade e que inclui como principais impulsionadores fabricantes de sistemas operativos, instituições públicas e privadas, militares e a indústria da segurança.

O SCAP inclui vários componentes; na versão 1.0 integra os seguintes:

- *Common Vulnerabilities and Exposures* (CVE) é certamente o mais difundido e foi adoptado pela generalidade da indústria das tecnologias de informação. Fornece métodos para identificação de vulnerabili-

>>>

dades e problemas de segurança de informação, os conhecidos CVE IDs incluídos nos anúncios de atualizações, como por exemplo o CVE-2013-1286 (referente a um problema do driver USB em sistemas Microsoft Windows) ou CVE-2013-0422 (referente a um problema no Security Manager no JAVA da Oracle). Estes identificadores são únicos, universais e podem ser aplicáveis simultaneamente a várias plataformas (como é o caso do referido problema no JAVA, que é transversal ao sistema operativo ou distribuição). Este formato foi definido pelo MITRE, que continua a ser a entidade responsável pela atribuição dos identificadores [1].

- *Common Configuration Enumeration* (CCE) é similar ao CVE mas refere-se a opções de configuração; da mesma forma define CCE IDs únicos que são aplicáveis a requisitos de configuração. Como exemplo, CCE-10583-3 (referente à necessidade de configurar a comunidade SNMP nos sistemas Windows 2008) ou CCE-4387-7 (referente à configuração que impede logins como utilizador root via SSH). Foi inicialmente definido pelo MITRE mas é agora mantido pelo NIST na *National Vulnerability Database* (NVD) [2].
- *Common Platform Enumeration* (CPE) define um método universal e normalizado para identificar e descrever aplicações, sistemas e dispositivos. A título de exemplo cpe:/a:microsoft:office:2007:sp2:professional refere-se ao "Microsoft Office 2007 Professional Service Pack 2" e cpe:/o:redhat:enterprise_linux:6 ao "Red Hat Enterprise Linux 6". Estes identificadores, assim como o formato propriamente dito são mantidos pelo NIST [3].
- *Common Vulnerability Scoring System* (CVSS) atribui níveis de severidade às vulnerabilidades consoante um conjunto de cenários e métricas. Permite indicar se um dado problema pode ser explorado através da rede (ou apenas localmente), qual o nível de complexidade do ataque, se requer autenticação prévia, qual o impacto que pode ter para o sistema (e sistemas adjacentes), etc. Apenas a título de exemplo, AV:L/AC:H/Au:N/C:C/I:C/A:C refere-se a uma vulnerabilidade apenas com impacto local, de utilização complexa, sem necessidade de autenticação, (...). Este protocolo é mantido pelo *Forum of Incident Response and Security Teams* (FIRST) [4].

Os protocolos até agora referidos pretendem apenas facilitar o processo de identificação e classificação de questões de segurança. Embora úteis na comunicação entre entidades, é difícil ver que vantagens trazem no que diz respeito à automatização. É aqui que entram os pesos-pesados do SCAP, os quais irão ser detalhados neste artigo mas que para já serão listados sem grande detalhe:

- *Open Vulnerability and Assessment Language* (OVAL) uniformiza a especificação de testes que permitem determinar a existência de recursos, opções de configuração, problemas de segurança, etc. As verificações são definidas por uma ou mais caraterísticas do sistema (ex. ficheiros, programas instalados, opções de configuração), qual o estado a testar para essa característica (conter o valor public ou estar instalado o pacote telnet-server) e qual o resultado que daí deverá extraído. Utiliza um formato XML definido e mantido pelo MITRE [5]; são vários os fabricantes (e outras entidades) a publicar testes OVAL, sendo cada vez mais habitual, por exemplo, a publicação das verificações OVAL em conjunto com a divulgação de vulnerabilidades ou patches.
- *eXtensible Configuration Checklist Description Format* (XCCDF) é um formato estruturado para especificação de *checklists, benchmarks* e documentação de segurança, central no SCAP. Este formato, também XML completa os restantes formatos SCAP com descrições textuais, procedimentos de remediação, etc. Este é o formato das políticas SCAP propriamente ditas e é mantido pelo NIST [6].

Versões posteriores do SCAP introduziram novos protocolos, os quais serão apenas resumidamente enumerados:

- *Asset Identification* (AI) para correlação de informação relativa a equipamentos e sistemas [7];
- *Open Checklist Interactive Language* (OCIL) define um sistema para a representação de questões interativas assim como os procedimentos para interpretar as respostas dadas pelos utilizadores; é corrente o conteúdo XCCDF incluir conteúdo OCIL, especialmente quando as verificações não são verificáveis automaticamente (ex. "o acesso físico ao sistema é restrito a pessoal autorizado") [8];
- *Asset Reporting Format* (ARF) define um formato para transferência de informação relativa a equipamentos e relatórios [9];
- *Common Configuration Scoring System* (CCSS) complementa o CVSS focando-se especificamente nas configurações [10];
- *Trust Model for Security Automation Data* (TMSAD) define um formato para validação e certificação dos vários elementos e documentos SCAP [11].

Outros protocolos são considerados emergentes, tentando abranger outros fluxos de informação e necessidades normativas [12].

Como se pode ver, o SCAP abrange um conjunto significativo de protocolos, e no âmbito deste artigo era impossível abordalos todos em maior detalhe.

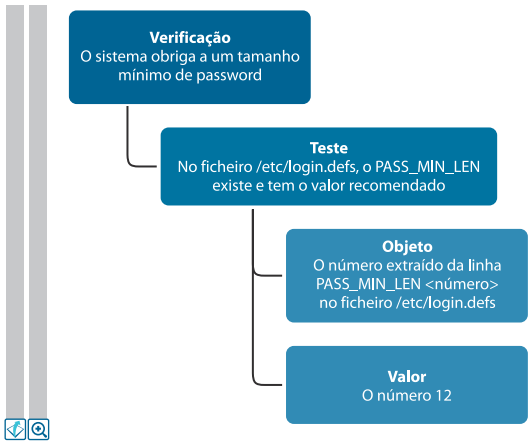
>>>

Será dado algum detalhe adicional aos protocolos OVAL e XCCDF por serem essenciais para os objetivos deste artigo.

2. DEFINIÇÃO DE TESTES (OVAL)

Duma forma muito resumida, pode-se dizer que o OVAL é um formato para definição de **verificações** como a conjugação lógica de vários **testes** que resultam da avaliação do estado dum **objeto** do sistema ou plataforma. A linguagem em si é baseada em XML mas a especificação formal fica para a norma.

Como exemplo simples, pretende-se verificar se um sistema Linux está em **conformidade** com a regra que obriga a que as *passwords* tenham pelo menos 12 caracteres. Sem entrar em considerações técnicas, o teste em Linux passa por verificar se existe no ficheiro /etc/login.defs uma entrada com "PASS_MIN_LEN 12":



Definida a verificação, é agora possível avaliá-la em qualquer sistema ou plataforma, mesmo onde o teste não faça qualquer sentido (como seria o caso de plataforma Windows). Apresenta-se o relatório resultante da avaliação num sistema Linux, depois de convertido para HTML para uma visualização simplificada:

Oval Definition Results

True | False | Error | Unknown | Not Applicable | Not Evaluated

OVAL ID	Result	Class	Reference ID	Title
oval:PTIN:def:1	true	compliance	Saber & Fazer	O sistema obriga a um tamanho mínimo de password

A **verificação** poderia ser mais elaborada: poderiam ser acrescentados outros testes (ex. validação com regras PAM – *Pluggable Authentication Modules*) ou incluídas outras plataformas (ex. incluindo o teste de forma a suportar Windows). Adicionalmente, as definições OVAL podem suportar variáveis externas; isto permite que, por exemplo, a mesma verificação possa ser reutilizada para postos de trabalho, onde o mínimo são apenas 8 caracteres.

No exemplo utiliza-se uma verificação do tipo conformidade. No entanto a norma define um total de cinco tipos de verificações:

- *vulnerability*: permite determinar se uma dada **vulnerabilidade** existe num dado sistema (ex. o sistema possui uma versão dum serviço com falhas de segurança conhecidas).
- *inventory:* permite determinar se um dado item (ex: pacote, versão de sistema operativo, etc.) está instalada, ou seja, pertence ao **inventário** do sistema.
- *patch*: permite verificar se uma determinada **correção** está aplicada no sistema; deverá sempre que possível conter a referência do problema que corrige.
- *compliance*: permite verificar se o sistema está em **conformidade** com regras de configuração (ex: tamanho mínimo da password).
- *miscellaneous:* para qualquer outro tipo de verificação.

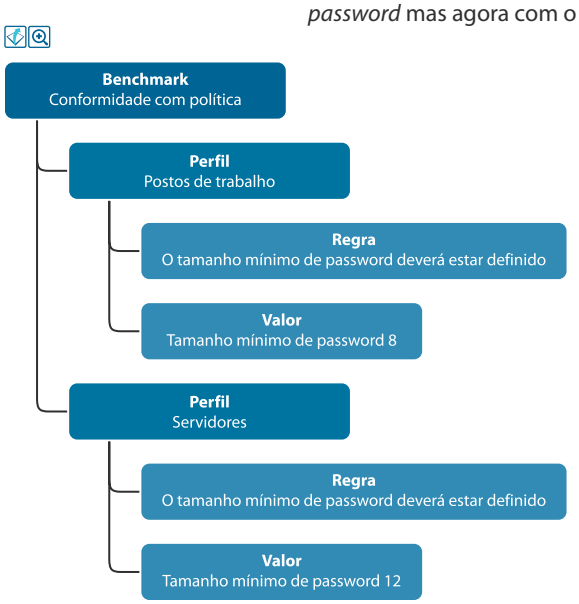
3. DEFINIÇÃO DE *BENCHMARKS* (XCCDF)

No ponto anterior viu-se como criar uma verificação, mas não foi dito em que contexto é que essa verificação pode ser aplicada. Foi também referido que a verificação pode suportar variáveis externas, mas não foi explicado como é que essas variáveis externas vão ser preenchidas e em que contexto nem com que valores.

>>>

O XCCDF vem endereçar estes pontos, pois permite definir ***benchmarks***, que contêm perfis, onde um **perfil** é nada mais que um conjunto de **regras** que relaciona as verificações OVAL (ou OCIL), as **referências** (CPE, CCE, etc.), **descrições** textuais, procedimentos de **remediação** (automáticos e manuais), etc.

Voltando ao exemplo, é possível ter um perfil "Servidores" com uma regra que relaciona a verificação anterior do tamanho mínimo da *password* com o valor 12. Adicionalmente, é possível criar um outro perfil "Postos de trabalho" com um regra que relaciona a mesma verificação do tamanho mínimo da *password* mas agora com o valor 8.



Os ficheiros XCCDF podem ser utilizados de várias formas. A mais corrente e útil é a avaliação de conformidade à política descrita. Por exemplo, no caso das *passwords*, o resultado da avaliação do perfil "Servidores" num sistema não conforme resultaria no seguinte relatório:

Score

system	score	max	%	bar
urn:xccdf:scoring:default	.00	100.00	.00%	

Results overview

Rule Results Summary

pass	fixed	fail	error	not selected	not checked	not applicable	informational	unknown	total
0	0	1	0	0	0	0	0	0	1

Title	Result
O tamanho mínimo de password deverá estar definido.	fail

Results details

Result for O tamanho mínimo de password deverá estar definido.

Result: **fail**

Rule ID: **password_min_len**

Time: **2013-08-20 11:59**

Os sistemas deverão ter configurados mecanismos que forcem um tamanho mínimo de password. Para este sistema, o tamanho mínimo aceitável é de 12 caractéres.

Garantir que o ficheiro `/etc/login.defs` contém:

```
PASS_MIN_LEN 12
```

O ficheiro login.defs tem PASS_MIN_LEN 10

path	content
/etc/login.defs	PASS_MIN_LEN 10

results overview

O relatório inclui exatamente qual a causa da verificação falhada (no exemplo, o sistema tem definido um tamanho mínimo de 10 carateres em vez de 12). O formato XCCDF pode incluir regras de **remediação automática**, via scripts, mas mesmo manualmente é trivial implementar as modificações necessárias para tornar o sistema conforme.

O resultado da avaliação do XCCDF é ele próprio um XML, que pode ser transformado para outros formatos que não o relatório HTML. Isto permite que sejam extraídos históricos e métricas da conformidade dos sistemas, permitindo por exemplo a geração de alarmes ou um cálculo de risco.

Além da avaliação da política, os ficheiros XCCDF podem dar origem a outros conteúdos:

- guias de configuração, que utilizam os campos descritivos, referências e conteúdos adicionais para produzir um documento em formato navegável e próximo do utilizado em políticas de segurança textuais;
- *checklists* de configuração utilizando o formato as referências CCE;
- qualquer outro output que resulte da transformação do XML do XCCDF.

4. UTILIZAÇÃO DO SCAP NA PTIN

Em 2002, e na sequência do colapso de companhias como a Enron e WorldCom, foi aprovado o **Sarbanes-Oxley Act**, ou S**OX** como é mais conhecido, que visou definir *standards* no comportamento das companhias publicamente cotadas nas bolsas dos EUA.

A PT é uma companhia cotada na NYSE, e por esse motivo está sujeita às mesmas regras. A PT Inovação, como empresa do Grupo PT e como fornecedor de sistemas para as restantes empresas do Grupo PT, tem como responsabilidade garantir que os sistemas que fornece seguem as políticas definidas pelo Grupo PT no geral e pela empresa cliente em particular.

O processo de instalação e configuração de uma solução num ambiente de cliente pode ser extremamente complexo, envolvendo vários servidores, cada um com centenas de pacotes de *software* desenvolvidos pela PT Inovação ou por terceiros, várias interligações com sistemas externos, diversos equipamentos de rede, *storage*, *backups*, etc. Em todo este processo, estão envolvidas várias pessoas quer da PT Inovação quer do cliente, a instalar, configurar e testar os sistemas para garantir que tudo está a funcionar devidamente.

Olhando apenas para a configuração do sistema operativo e *software* infra-estrutural como os servidores aplicacionais, as bases de dados, etc; as configurações necessárias podem ser diferentes de cliente para cliente, devido às diferentes políticas de segurança de cada cliente. Garantir que todos os servidores e sistemas estão devidamente configurados e seguem as políticas definidas é uma tarefa extremamente complexa se efetuada manualmente.

A PT Inovação tem vindo a efetuar uma migração gradual dos seu produtos de Linux para a utilização do formato RPM. A utilização de RPMs para instalação de *software* permite automatizar este processo normalmente complexo e propício a erros, garantindo determinismo, integridade do *software*, *rollbacks* automatizados, gestão de dependências, estandardização da forma de instalar e operar os sistemas bem como da localização de ficheiros de configurações, logs, bibliotecas, etc. Adicionalmente, passa a ser trivial identificar que pacotes de *software* e que versões estão instaladas num determinado sistema. Este é um passo fundamental para todo o processo de automatização descrito abaixo.

5. CRIAÇÃO DE PERFIS DE *BENCHMARK*

Impulsionado pela Red Hat, o projeto SCAP *Security Guide* (SSG [13]) contém um grande número de verificações OVAL e conteúdo XCCDF baseados nas melhores práticas e em normas como os *Security Technical mplementation Guides* (STIGs [14]) e o *United States Government Configuration Baseline* (USGCB [15]), respetivamente da *Field Security Operations* (FSO) da *Defense Information Systems Agency* (DISA) e do próprio NIST.

O conteúdo SSG, além de constituir uma boa base, introduz facilidades para edição por equipas e algumas simplificações ao formato SCAP, pelo que foi escolhido pela PT Inovação para o desenvolvimento dos vários perfis de *benchmark* para sistemas baseados em Red Hat Entreprise Linux, nomeadamente:

- production-base, cujo objetivo é estabelecer um mínimo absoluto e obrigatório com 100% de conformidade para todos os sistemas de produção, tendo como base a Política de Segurança da Informação da PT Portugal e algumas boas práticas consensuais;
- production-services, que expande o perfil base com regras adicionais especialmente focadas em serviços, cuja principal função é garantir que apenas estão ativos os serviços mesmo necessários para a solução;

>>>

- client-profile-n, (uma ou várias por cliente) derivados também do perfil base com a adaptação de regras específicas dos clientes; é obrigatória a conformidade a 100% em sistemas de produção destinados ao cliente. Estes perfis são construídos de acordo com as políticas dos clientes.

Estes perfis estão disponíveis num repositório local com conteúdo SCAP, que é atualizado com uma réplica diária do ficheiro OVAL disponibilizado pela Red Hat com informação de atualizações. Estes perfis são também atualizados sempre que necessário para irem de encontro a novas regras ou verificações que sejam identificadas pela própria PT Inovação ou clientes.

6. FERRAMENTA DE AUTOMATIZAÇÃO

Tendo estes perfis definidos, foi necessário implementar uma forma que facilite a configuração inicial de um servidor de acordo com os perfis definidos, bem como a forma de verificar regularmente que o servidor continua de acordo com os perfis definidos, principalmente no que diz respeito a atualizações de segurança.

Para efetuar a configuração inicial estão implementados vários scripts, empacotados sob a forma de RPMs, que ao serem instalados executam automaticamente as configurações identificadas no respetivo perfil. Com estes pacotes, e baseando a instalação de servidores em *templates* do *kickstart* da Red Hat pré definidos, é garantido que a instalação inicial do servidor está conforme a políticas de segurança definidas, e com uma configuração uniforme e bem conhecida, evitando o problema conhecido dos "Snowflake servers" [16].

Adicionalmente, e usando por base as ferramentas disponibilizadas pelo projeto Open SCAP disponível em várias versões UNIX e com suporte para a versão 1.2 do SCAP (ainda que incompleto), foi criada uma ferramenta check-security, distribuída igualmente por RPM e cujas principais funcionalidades são:

- atualiza automaticamente os conteúdos SCAP através do acesso a repositórios internos PT Inovação, e com suporte para a criação e utilização de outros repositórios existentes numa rede local; a ideia será manter e propagar periodicamente os repositórios SCAP da PT Inovação para um servidor disponível na rede do cliente, permitindo assim verificação de todas as máquinas existentes no cliente sem dependência do acesso à rede da PT Inovação;
- verifica o estado dos *patches* do sistema (via OVAL disponibilizado pela Red Hat), o que permite identificar falhas de segurança conhecidas e qual a sua criticidade;
- testa a conformidade em relação aos perfis instalados, permitindo identificar mudanças de configuração que não estejam de acordo com as definições dos perfis;
- em caso de falha de ligação a um servidor com conteúdo SCAP atualizado, funciona com o conteúdo por omissão;
- por fim, gera automaticamente os relatórios de conformidade em vários formatos (incluindo HTML), que podem ser analisados ou arquivados para histórico.

Esta ferramenta pode ser utilizada interativamente ou integrada em verificações programadas do sistema. A avaliação de conformidade e *patches* tem de ser executada regularmente, com conteúdos atualizados, em todos os servidores de produção, sendo recomendado uma periodicidade mínima de uma semana para servidores acessíveis a partir da Internet.

Sem nos detalhar todas as opções de funcionamento, no seu caso mais simples a execução é:

```
# check-security
Getting content updates
--> Fetching PTIN SCAP definitions from http://pds.ptin.corppt.com/scap ...
Success.
--> Fetching RHEL OVAL definitions from http://pds.ptin.corppt.com/scap/com.
redhat.rhsa-all.xml.bz2 ... Success.
Checking Redhat Patch Compliance...
--> Evaluating OVAL 'com.redhat.rhsa-all.xml'... system FAILS 2 security updates
Checking PTIN Policies...
--> Evaluating XCCDF 'production-base'... system FAILS 13 checks.
#
```

Isto irá resultar (entre outros) em dois ficheiros HTML, contendo o resultado das avaliações no sistema:

Verificação de Atualizações (RHEL OVAL):

OVAL Definition Results

Error | Unknown | Other

ID	Result	Class	Reference ID	Title
oval:com.redhat.rhsa:def:20131459	true	patch	[RHSA-2013:1459-00], [CVE-2012-6085], [CVE-2013-4351], [CVE-2013-4402]	RHSA-2013:1459: gnupg2 security update (Moderate)
oval:com.redhat.rhsa:def:20131457	true	patch	[RHSA-2013:1457-00], [CVE-2013-4242]	RHSA-2013:1457: libgcrypt security update (Moderate)
oval:com.redhat.rhsa:def:20131436	true	patch	[RHSA-2013:1436-00], [CVE-2013-4162], [CVE-2013-4299]	RHSA-2013:1436: kernel security and bug fix update (Moderate)
oval:com.redhat.rhsa:def:20131458	false	patch	[RHSA-2013:1458-00], [CVE-2012-6085], [CVE-2013-4242], [CVE-2013-4351], [CVE-2013-4402]	RHSA-2013:1458: gnupg security update (Moderate)

Este conteúdo cruza as *Red Hat Security Advisories* (RHSA) com os respetivos CVE-ID, resultando numa imagem completa sobre quais os pacotes com problemas que estão instalados no sistema, assinalados no relatório com a cor laranja e indicado o seu grau de criticidade. O relatório é um resultado abreviado do conteúdo da Red Hat, que contém bastante mais informação e referências; no entanto este demonstra ser suficiente para uma utilização prática.

Verificação de Conformidade (PTIN SCAP):

Score

system	score	max	%	bar
urn:xccdf:scoring:default	88.44	100.00	88.44%	

Results overview

Rule Results Summary

pass	fixed	fail	error	not selected	not checked	not applicable	informational	unknown	total
104	0	13	0	283	0	0	0	0	400

Title	Result
Ensure /tmp Located On Separate Partition	fail
Ensure /var/log Located On Separate Partition	pass
Ensure /home Located On Separate Partition	pass
Ensure Red Hat GPG Key Installed	pass
Verify File Hashes with RPM	fail
Add nodev Option to Non-Root Local Partitions	pass
Add nodev Option to /tmp	fail
Add nosuid Option to /tmp	fail
Add nodev Option to /dev/shm	pass
Add noexec Option to /dev/shm	pass
Add nosuid Option to /dev/shm	pass
Disable the Automounter	pass
Verify User Who Owns shadow File	pass
Verify Group Who Owns shadow File	pass
Verify Permissions on shadow File	pass
Verify User Who Owns group File	pass
Verify Group Who Owns group File	pass
Verify Permissions on group File	pass
Verify User Who Owns gshadow File	pass
Verify Group Who Owns gshadow File	pass

Além do sumário do relatório, é possível verificar como foi avaliada a conformidade para cada uma das regras. Por exemplo, o sistema avaliado tem instalado o pacote telnet-server consensualmente inseguro:

Score

system	score	max	%	bar
urn:xccdf:scoring:default	.00	100.00	.00%	

Results overview

Rule Results Summary

pass	fixed	fail	error	not selected	not checked	not applicable	informational	unknown	total
0	0	1	0	0	0	0	0	0	1

Title	Result
O tamanho mínimo de password deverá estar definido.	fail

Results details

Result for O tamanho mínimo de password deverá estar definido.

Result: fail

Rule ID: **password_min_len**

Time: **2013-08-20 11:59**

Os sistemas deverão ter configurados mecanismos que forcem um tamanho mínimo de password. Para este sistema, o tamanho mínimo aceitável é de 12 caractéres.

Garantir que o ficheiro `/etc/login.defs` contém:

```
PASS_MIN_LEN 12
```

O ficheiro login.defs tem PASS_MIN_LEN 10

path	content
/etc/login.defs	PASS_MIN_LEN 10

results overview

Além das instruções de remediação manual, é incluída a informação de remediação automática ("*Remediation Script*") que poderia ser executada diretamente pelo administrador de sistema. O Open SCAP pode automaticamente executar as *scripts* de remediação para as regras que falharam (e para as quais tenha definidas regras de remediação).

Ou seja, enquanto os RPMS com os scripts de configuração garantem que o servidor nasce **seguro após a instalação**, a utilização regular desta ferramenta vai garantir que o servidor se **mantém seguro ao longo do tempo**, e que as mexidas do dia-a-dia não estragam essa *baseline*.

Tendo a solução técnica definida, é então necessário garantir que o elo fraco em todo este processo da segurança, e que são as pessoas, não deitam tudo a perder. Mesmo estando bem -intencionadas, as pessoas por vezes seguem por "atalhos" que colocam em causa todo o trabalho feito até aí. Para isso, foi adicionada a "Instrução – Regras de Segurança" ao Processo de Desenvolvimento de Sistemas (PDS) que documenta o processo e as regras a cumprir por toda a PT Inovação, aliado a um esforço interno de divulgação e sensibilização, tentando motivar e convencer os colaboradores em geral para estes assuntos.

7. CONCLUSÃO E TRABALHOS FUTUROS

Este é um processo que está ainda na sua fase inicial, no entanto começam já a ser visíveis as vantagens. Para isso, basta olhar para os primeiros resultados práticos que foram verificados:

- as equipas de instalação passaram a ter ao dispor procedimentos de instalação alinhados com as políticas de segurança, o que além dos ganhos em produtividade conseguidos pelo automatismo, permite o fornecimento de sistemas em conformidade com as melhores regras de segurança;
- os clientes podem agora comparar os resultados da validação dos perfis aos seus próprios requisitos, dispensando assim certificações e verificações morosas e dispendiosas;

- os administradores de sistemas dispõem duma ferramenta simples que ao ser executada regularmente pode detetar modificações no sistema; como exemplo, uma das regras da política base verifica se houve modificação nos ficheiros de sistema, vetor comum de ataque e que assim pode ser facilmente detetado;
- os auditores de segurança podem agora ter uma imagem instantânea do nível de segurança dos sistemas, tanto no que diz respeito a conformidade com as políticas como com o estado das atualizações do sistema;
- o fornecimento dum produto uniformizado no que diz respeito a configurações de segurança, mesmo quando o destinatário não dispõe de políticas próprias, é uma mais valia clara e permite responder a pressões por parte dos clientes no sentido de sistemas mais seguros e alinhados a normas.

Tendo alcançado esta primeira meta, o objetivo agora é estender o âmbito das regras e verificações definidas. Neste momento ainda não existem regras próprias para muitos dos produtos de *middleware* de que dependem os sistemas desenvolvidos pela PT Inovação (ex: Bases de Dados, Servidores Aplicacionais, etc.), e é necessário endereçar esses componentes, pois muitos dos problemas de segurança acontecem exatamente neste ponto.

Este artigo foi muito focado em Red Hat Linux, porque esta é a base tecnológica da grande maioria das soluções desenvolvidas pela PT Inovação. No entanto é objetivo envolver outras plataformas e equipas, uma vez que o SCAP é utilizado noutros sistemas: por exemplo, a Microsoft integra capacidades SCAP no System Center Configuration Manager; a Cisco disponibiliza conteúdo OVAL para validar configurações e versões de produtos; a McAfee e a Symantec utilizam OVAL para definição e distribuição de políticas de conformidade; etc.

Embora seja impossível resolver definitivamente todos os problemas de segurança duma forma tecnológica, as políticas e em particular a conformidade são o caminho a seguir. Foi objetivo deste artigo transmitir a importância do SCAP como um passo na direção certa, bem como o aumento de eficácia e eficiência que se consegue nas equipas de instalação e operação.

REFERÊNCIAS

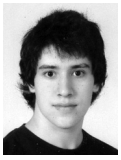
[1] http://cve.mitre.org
[2] http://nvd.nist.gov/cce.cfm
[3] http://nvd.nist.gov/cpe.cfm
[4] http://www.first.org/cvss
[5] http://oval.mitre.org/
[6] http://scap.nist.gov/specifications/xccdf/
[7] http://scap.nist.gov/specifications/ai/
[8] http://scap.nist.gov/specifications/ocil/
[9] http://scap.nist.gov/specifications/arf/
[10]http://csrc.nist.gov/publications/PubsNISTIRs.html#NIST-IR-7502
[11] http://scap.nist.gov/specifications/tmsad/
[12] http://scap.nist.gov/emerging-specs/listing.html.
[13] https://fedorahosted.org/scap-security-guide/
[14] http://iase.disa.mil/stigs/
[15] http://usgcb.nist.gov/
[16] http://martinfowler.com/bliki/SnowflakeServer.html

CVS DOS AUTORES

Mário Moreira obteve o mestrado em Ciências da Computação pela Universidade do Minho em 1995 e a Licenciatura em Engenharia de Sistemas e Informática pela Universidade do Minho em 1993. Ingressou no Centro de Estudos de Telecomunicações (futura PT Inovação), ainda durante a licenciatura em 1993 onde esteve a desenvolver vários sistemas periciais baseados em técnicas de Inteligência Artificial. Esteve inicialmente envolvido em vários projetos europeus de investigação ACTS e Eurescom. Desempenhou funções na área de serviços de gestão de redes, tendo colaborado com a equipa da Portugal Telecom no âmbito da Expo98, e estando envolvido no desenvolvimento de vários produtos da PT Inovação nesta área. Em 2001 ficou responsável pela Unidade de Plataformas, Serviços e Aplicações para Redes Móveis, onde esteve envolvido no desenvolvimento de várias plataformas de Unified Messaging, SMS-C, Localização e Portais Móveis. Atualmente é Gestor da Divisão de Tecnologias e Desenvolvimento na direção de Coordenação Tecnológica e Inovação Exploratória.

Rui Pedro Bernardino, licenciado em Engenharia de Sistemas e Informática da Universidade do Minho, iniciou a vida profissional no então CET em 1993 com estágio em redes inteligentes. Depois duma curta passagem pelo IT, foi em 1995 para a Parque Expo'98 onde foi responsável por vários sistemas de segurança antes e durante a exposição. Várias telcos e instituições financeiras depois, regressou à PTIN para desenvolver o projeto de segurança no âmbito do CTE.

07 Reforço da Privacidade Através do Controlo da Pegada Digital



RICARDO AZEVEDO RICARDO MACEDO

PALAVRAS CHAVE
"Sticky policies", sistema criptográfico, IBE, controlo de acessos, RBAC, linguagem de políticas, auditoria de ficheiros

A relação entre utilizadores e fornecedores de serviços é, por norma, assimétrica. A prática comum é, aquando da inscrição ou contratação de um serviço, o utilizador aceitar um conjunto de políticas referentes ao uso de informação privada facultada (ex., a morada, o número de telefone, preferências, etc...). Geralmente os utilizadores aceitam a política do fornecedor com base na confiança, na reputação, ou por qualquer outro fator (ex. recomendação), mas quase nunca com base numa leitura, e compreensão, apropriada das condições aplicáveis. Os casos de violação de privacidade por parte de alguns fornecedores, que vendem ou oferecem informação privada sobre os seus clientes a outros, são amplamente conhecidos e resultam em grande medida da falta de controlo que os utilizadores finais têm sobre a informação que disponibilizam para poder utilizar os diferentes serviços digitais. O problema enunciado não é novo, não se limitando ao espaço digital, mas a sua dimensão deve-se à proliferação de serviços e de informação privada que os utilizadores disponibilizam para os usar.

Também o ambiente empresarial está sujeito a este problema, ou variantes. Quase toda a informação de uma organização é guardada em claro, e usada por um variado conjunto de serviços e utilizadores (i.e. colaboradores da empresa). Mesmo que a informação seja guardada num local seguro, os administradores podem ter, indevidamente, acesso a informação privada da organização. Além disso, a organização pode ser alvo de ataques internos ou externos, que bem-sucedidos poderão garantir o acesso a dados importantes. É, aliás, comum que a comunicação, quando relativo a decisões estratégicas e que se pretendem secretas, entre administradores da mesma empresa seja feita utilizando servidores de *email* externos à empresa.

Neste trabalho propomos a implementação de um mecanismo que possibilite o envio de informações sem que o utilizador tenha necessidade de se preocupar com o nível de confiança do local onde as mesmas serão armazenadas, através da utilização do conceito de "sticky policies". Através da utilização de técnicas criptográficas, é estabelecido um vínculo entre a informação cifrada e as políticas de autorização que regem o acesso à informação. O sistema desenvolvido garante que, para um terceiro aceder às informações pessoais de um utilizador, terá que cumprir o conjunto de regras definidas pelo dono da informação. O processo de autorização, passa a ser um processo distribuído – e não executado pelo serviço –, uma vez que as regras de autorização estão *coladas* à própria informação.

Visto que um utilizador autorizado a aceder às informações pode ter um comportamento incorreto, partilhando indevidamente as informações, propomos também adicionar mecanismos de auditoria dos acessos à informação gerida pelo sistema.

1. INTRODUÇÃO

O crescimento da *internet* e da informatização dos serviços, que se tem vindo a testemunhar, promoveu inúmeras vantagens para a população, mas com elas têm surgido algumas questões relacionadas com a segurança e a privacidade de dados privados.

Um dos grandes problemas da utilização da *internet* é o risco de desconhecidos terem acesso a informações pessoais. Isto acontece quando os utilizadores são alvo de ataques cibernéticos (*hacking*) ou quando são divulgadas indevidamente por utilizadores autorizados. Depois de informações pessoais serem enviadas a um terceiro, este pode-as usar incorretamente ameaçando a privacidade do utilizador.

Devido a este problema, quando um utilizador envia informações pessoais necessita de confiar no destinatário. Esta necessidade é responsável pelo aparecimento de dois tipos de utilizadores, aqueles que confiam em tudo, disponibilizando constantemente as suas informações pessoais, e por outro lado, os que não confiam em nada, privando-se assim da utilização de muitos serviços.

2. ENQUADRAMENTO E MOTIVAÇÃO

No seio das empresas existe também este problema de partilha de informação. Como as informações são armazenadas em claro, ainda que estejam num local seguro, todos os funcionários que tiverem acesso à base de dado têm acesso a todas as informações lá armazenadas. E na situação de a empresa ser alvo de um ataque informático, se o atacante conseguir aceder à base de dados pode consultar livremente todas as informações.

Alan Giang Tran, em janeiro de 2003, efetuou um ataque informático aos computadores das empresas Airline Coach Service e Sky Limousine (da mesma sociedade), onde trabalhou como administrador de redes. No ataque foram alteradas todas as *passwords* do sistema, resultando em grandes prejuízos nas empresas [1]. Neste ataque, o atacante em vez de trocar as *passwords* poderia ter acedido às informações internas, para posteriormente as vender a outras empresas, resultando em ainda maiores prejuízos.

Neste projeto foi criado um sistema de partilha de ficheiros em que o utilizador controla quem pode aceder aos ficheiros. Neste sistema a informação pode ser armazenada em qualquer lugar, visto que é protegida antes de ser armazenada. O sistema é funcionalmente modular e permite que se possa integrar com *Content Management Systems* empresariais.

3. CENÁRIO

O cenário de demonstração escolhido consiste num sistema de partilha de ficheiros que realizará o controlo de acessos baseado nos papéis (*roles*) dos trabalhadores, no contexto da empresa.

Para o sistema realizar o controlo de acessos, quando um utilizador faz o *upload* de um ficheiro deverá descriminar que papéis deverão ter os utilizadores para poderem aceder ao ficheiro.

Para armazenar os ficheiros será utilizado o serviço de armazenamento, que pode ser um qualquer "Cloud provider", para isto, é preciso considerar que o serviço de armazenamento não será seguro, podendo estar sujeito a ataques, ou alterações unilaterais de políticas do serviço.

>>>

Para se armazenar ficheiros num local inseguro, é necessário protege-los antes de os armazenar, para isso é utilizado um serviço criptográfico, que é capaz de cifrar e decifrar ficheiros.

Além disto, também é importante que seja viável a utilização do sistema em qualquer dispositivo, independentemente da sua capacidade computacional e autonomia, dai que tenhamos optado por ter o sistema criptográfico exposto como um serviço.

Como se vê na figura 1, o utilizador "*User 1*" envia para o sistema um ficheiro e especifica que apenas pode ser acedido pelos utilizadores com papel "admin". O sistema ao receber o ficheiro invoca o serviço criptográfico, de modo que o ficheiro seja cifrado criando uma ligação entre o criptograma e as políticas definidas. De seguida o sistema envia o ficheiro cifrado e as políticas associadas para o serviço de armazenamento. Quando um utilizador pretende aceder ao ficheiro o sistema recupera o ficheiro e as suas políticas de autorização do serviço de armazenamento e recorre ao serviço criptográfico enviando-lhe o criptograma, as politicas associadas e a identificação do utilizador que está autenticado no serviço. Se o utilizador tiver permissões de acesso ao ficheiro, segundo as regras de autorização estabelecidas anteriormente, o ficheiro é decifrado. É assumido que toda a comunicação entre os intervenientes é executada sobre um canal seguro (i.e. HTTPS).

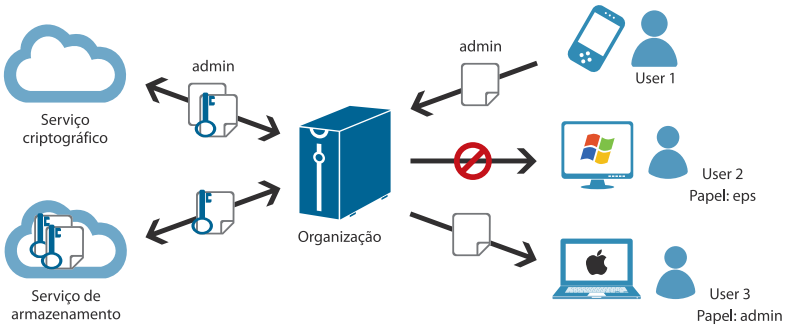


Figura 1. Esquema do cenário de demonstração

4. DESENVOLVIMENTO

O sistema implementado, "Sticky-Policies File Sharing" [2][1], segue o esquema apresentado na figura 2. Como o foco do projeto era a arquitetura de "sticky policies", optou-se por não se implementar o serviço de armazenamento, armazenando os ficheiros localmente no disco. Ainda na figura 2, podemos ver que o sistema utiliza o IAM para autenticar os utilizadores.

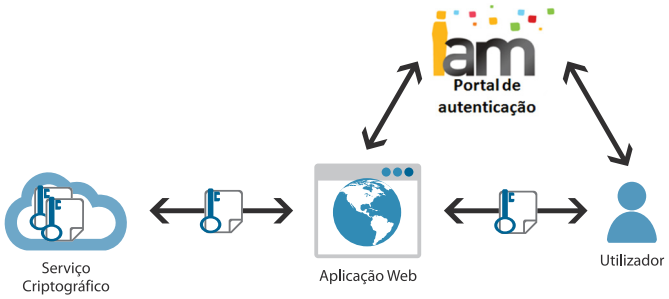


Figura 2. Esquema do sistema implementado

O esquema criptográfico utilizado foi o "Identity Based-Encryption" (IBE) [3], este esquema permite cifrar informações usando uma identidade como chave pública. Para se criar a ligação entre criptograma e o ficheiro com a política de controlo de acesso, usou-se como chave pública o "hash" gerado com o ficheiro da política. Para implementar os algoritmos do esquema IBE em Java foi utilizada a

[1] Pode experimentar o serviço, autenticando-se com as suas credenciais PTIN.

biblioteca jPBC [4]. Apesar de se utilizar o esquema IBE, a informação não é cifrada com este esquema, tal como os outros esquemas criptográficos assimétricos, também no esquema IBE os tempos de execução são geralmente longos. Para resolver este problema foi utilizado o procedimento habitual de utilizar uma cifra simétrica para cifrar a informação, pois as cifras simétricas têm tempos de execução muito menores. Assim para cifrar um ficheiro é gerada aleatoriamente uma chave simétrica que é utilizada para cifrar o ficheiro, e de seguida é utilizado o esquema IBE para cifrar a chave simétrica utilizada. Para se especificar as políticas de controlo de acessos (autorização) foi escolhida a linguagem de políticas XACML [5] e para desenvolver o sistema foi utilizada a implementação XACML da Sun [6]. Note-se que os utilizadores não têm de conhecer a linguagem. Optou-se por utilizar uma linguagem de definição de política já existente, e considerada *standard*, em vez de criar uma nova forma de escrever política de autorização.

O serviço criptográfico segue a arquitetura presente na figura 3. É composto por quatro unidades funcionais, o "Service", o "Encryptor", o "Decryptor" e o "Keyman", que comunicam entre si, como está representado na figura 3.

O "Service" é o componente que comunica com o exterior, recebe os pedidos de *upload* ou *download* de um ficheiro e para estes pedidos é responsável por fazer o controlo das políticas de controlo de acessos.

Por sua vez, o "Keyman" é o responsável por gerar os parâmetros públicos e as chaves privadas do esquema IBE. Quando este recebe pedidos para gerar chaves privadas, apenas responde quando os pedidos são feitos pelo "Decryptor", caso contrário para um atacante ser capaz decifrar um ficheiro apenas necessitava de realizar um pedido de geração da chave privada ao "Keyman".

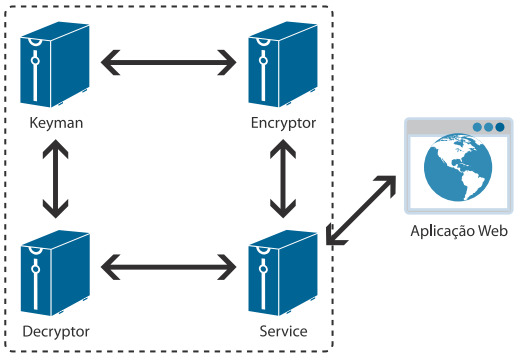


Figura 3. Arquitetura do serviço criptográfico

Por fim, o "Encryptor" e o "Decryptor" são responsáveis por cifrar e decifrar ficheiros, respetivamente. O "Decryptor" é ainda responsável por validar a integridade do ficheiro e de validar se as regras de autorização associadas ao ficheiro permitem que o utilizador, que está a fazer o pedido, tenha acesso ao ficheiro.

5. RESULTADOS

Para se testar o sistema foi utilizada uma máquina com Windows *7 Enterprise 64 bits*, com um processador Intel Core i5-3210M e 4096 MB de RAM.

Na tabela 1 é possível ver os tempos de execução dos algoritmos da cifra simétrica (cifrar o ficheiro) e assimétrica (cifrar a chave simétrica utilizada para cifrar o ficheiro) no *upload* de um ficheiro. Cada linha representa o *upload* de um ficheiro. A primeira coluna apresenta o tamanho do ficheiro, a segunda o tempo de cifra do ficheiro - através da cifra simétrica - e a terceira coluna apresenta o tempo de cifra da chave simétrica utilizando o esquema IBE.

Tamanho do Ficheiro	Tempo de Cifrar o Ficheiro	Tempo de Cifrar a Chave
14,38 KB	7 ms	139 ms
67,34 KB	13 ms	151 ms
272 KB	27 ms	144 ms
696,51 KB	30 ms	138 ms
1,66 MB	36 ms	142 ms
18,66 MB	795 ms	139 ms

Tabela 1. Tempos de execução no *upload* de um ficheiro

Relativamente à tabela 2, são apresentados os tempos de execução dos algoritmos da cifra assimétrica (decifrar a chave simétrica com a qual foi cifrado o ficheiro) e simétrica (decifrar o ficheiro) no *download* de um ficheiro.

Tamanho do Ficheiro	Tempo de Decifrar a Chave	Tempo de Decifrar o Ficheiro
14,38 KB	70 ms	4 ms
67,34 KB	71 ms	17 ms
272 KB	71 ms	18 ms
696,51 KB	67 ms	15 ms
1,66 MB	69 ms	39 ms
18,66 MB	68 ms	869 ms

Tabela 2. Tempos de execução no *upload* de um ficheiro

Por fim, na tabela 3 são apresentados os tempos de execução do algoritmo de geração de chaves privadas do esquema IBE, em que cada linha representa uma chave gerada. Como este algoritmo apenas depende do ficheiro da política de controlo de acessos e não do ficheiro a decifrar, na primeira coluna é apresentado o número de papéis que estão autorizados pela política e na segunda o tempo que levou a ser gerada a chave privada.

Número de Papéis	Tempo de Gerar a Chave Privada
1	95 ms
2	94 ms
3	93 ms
4	94 ms

Tabela 3. Tempos de execução de gerar chaves privadas do esquema IBE

Através da análise das tabelas podemos concluir que o tempo de execução dos algoritmos do esquema IBE são desprezáveis, isto é, no contexto de um *upload* ou o *download* de um ficheiro ter um acréscimo 150 milissegundos no *upload* e um acréscimo de 165 milissegundos no *download* (tempo de gerar a chave privada mais o de decifrar a chave simétrica) são desprezáveis.

Os tempos de execução dos algoritmos do esquema IBE mantêm-se constantes com o crescimento dos ficheiros e com o crescimento das políticas. No que respeita ao tempo de cifrar e decifrar a chave simétrica mantém-se constante pois o tamanho da chave simétrica é independente do tamanho do ficheiro, permanecendo constante mesmo para ficheiros de tamanhos diferentes. O tempo de gerar a chave privada do esquema IBE também se mantém constante com o crescimento das políticas pois o que é utilizado como chave pública não é a política em si, mas um "hash" gerado através dessa política, assim o tamanho do "hash" também é constante.

6. CONCLUSÕES E TRABALHO FUTURO

Este projeto vem resolver o problema de a informação partilhada nas empresas ser armazenada em claro, podendo assim ser acedida indevidamente. Para isto foi desenvolvido o sistema apresentado que através de um conjunto de regras de autorização, definidas pelo dono da informação, realiza o controlo de acessos aos ficheiros.

Além de garantir que o acesso aos ficheiros é feito com base nas políticas de autorização definidas, o sistema garante, também a integridade dos ficheiros, isto é, qualquer alteração maliciosa é detetada no momento em que se decifra o documento.

Os tempos de execução dos algoritmos utilizados no esquema IBE são baixos e constantes, podem por isso, ser desprezáveis. Para o tempo de resposta do sistema, o que tem maior influência são os tempos de execução dos algoritmos criptográficos simétricos que são utilizados para cifrar e decifrar os ficheiros. Estes aumentam com o aumento do tamanho dos ficheiros, mas ainda assim são tempos que permitem um bom funcionamento do sistema.

Propomos que futuramente seja usado um sistema de *Cloud* como serviço de armazenamento, em vez de se armazenar os ficheiros localmente.

REFERÊNCIAS

[1] http://www.justice.gov/criminal/cybercrime/press-release/2003/tranarrest.htm/

[2] https://sso-dev-evl-iam.ptin.corppt.com/spfilesharing-web/

[3] Dan Boneh and Matt Franklin. Indentity-based encryption from the weil pairing. In Joe Kilian, editor, *Advances in Cryptology – CRYPTO 2001*, volume 2139 of *Lecture Notes in Computer Science*, pages 213-229. Springer Berlin Heidelberg, 2001.

[4] http://gas.dia.unisa.it/projects/jpbc/

[5] The OASIS technical committee. eXtensible Access Control Markup Language (XACML) Version 2.0. 2005.

[6] http://sunxacml.sourceforge.net/

CVS DOS AUTORES

Ricardo Azevedo, MSC em Internet Computing pelo Queen Mary College (Universidade de Londres), em 2006, e Licenciado em Engenharia de Computadores e Telemática pela Universidade de Aveiro, em 2004. Foi bolseiro de investigação do IT desde 2004 a Fevereiro de 2006, com trabalho desenvolvido na área de QoS em redes heterogéneas de 4ª geração. Em Fevereiro de 2006 iniciou a sua actividade profissional na PT Inovação. Tem participado em diversos projectos de I&D no âmbito do IST e Eurescom, nas áreas de QoS e Network Management, Mobilidade, Segurança, Privacidade e Gestão de Identidades. Atualmente é Team Leader do grupo de Gestão de Identidades e Privacidade.

Ricardo Joaquim Pereira de Macedo, obteve o grau de licenciado e mestre em Engenharia Informática na Universidade do Minho, nos anos de 2011 e de 2013, respetivamente. Entrou na PT Inovação em fevereiro de 2013 como estagiário onde desenvolveu a sua dissertação de mestrado que deu origem ao presente artigo. Atualmente encontra-se a trabalhar na área de autenticação, autorização e controlo de acessos em big data no âmbito dos projetos europeus SMARTIE, UCN e CaaS.

08 Identificação Biométrica e Comportamental de Utilizadores em Cenários de Intrusão

HENRIQUE MARTINS RICARDO AZEVEDO

PALAVRAS CHAVE
Identificação, Autenticação, Segurança, Deteção de Intrusões, Biometrias Comportamentais, *Keystroke Dynamics*, *Mouse Dynamics*

Um sistema de autenticação forte é crucial. A confidencialidade, controlo de acesso e proteção contra ameaças internas e externas de recursos confidenciais são exigências atuais. Todos os métodos de autenticação existentes apresentam problemas, não garantindo, de forma completa, a identidade de um utilizador que usou um conjunto correto de credenciais. As palavras-chave podem ser esquecidas ou perdidas, os cartões podem ser duplicados ou roubados, e as biometrias físicas, também copiáveis, são intrusivas e a sua utilização obriga à compra de equipamento dispendioso. A utilização simultânea de vários métodos permite aumentar o nível de certeza de que o utilizador é quem diz ser, reduzindo para valores marginais os problemas apontados aos métodos usados habitualmente. É por isso que cada vez mais serviços, nomeadamente na área financeira, requerem a utilização de *tokens* para a realização de determinadas operações.

A autenticação inicial, num serviço, é crítica, mas as ameaças não se limitam apenas a esta fase. O intervalo entre o início e fecho de sessão não é monitorizado, ou seja, é assumido que o utilizador autenticado é aquele que usa a sessão durante todo o tempo. Mas isso pode não ser verdade.

Neste artigo é apresentada uma solução para o reforço aos atuais sistemas de autenticação e uma solução para a deteção de intrusões (mecanismo de autenticação contínua). A solução, para ambos os cenários, assenta na criação e validação de perfis comportamentais do utilizador. Este sistema não tem custos, já que é apenas usado equipamento básico, sendo completamente invisível para o utilizador. O protótipo desenvolvido foi incorporado no produto I*dentity and Access Manager (IAM)* da PT Inovação, fazendo parte dos módulos disponíveis para validação da autenticação.

1. INTRODUÇÃO

A usurpação de contas e o roubo de identidade são problemas muito frequentes nos atuais sistemas informáticos. A facilidade de acesso à internet e a exposição das pessoas a este meio, torna muito frequente a utilização indevida e a usurpação de contas (tais como: e-mail, redes sociais, contas bancárias) por outras pessoas que não as suas legítimas proprietárias.

Atualmente o método de autenticação dominante é o da combinação nome de utilizador e palavra-chave. No entanto, este método pode não ser fiável, pois estas credenciais podem ser partilhadas, roubadas ou até esquecidas [1]. A utilização de outros métodos de autenticação, geralmente aceites como mais seguros, têm vindo a ser usados, resultando num reforço efetivo da segurança dos sistemas [1]. Cartões de acesso, certificados digitais e biometrias são alguns dos métodos. No entanto todos são passíveis de fraude. Os cartões de acesso, como por exemplo os das caixas multibanco, podem ser roubados ou duplicados, como é frequentemente noticiado. Os certificados podem também ser copiados, distribuídos por correio eletrónico ou em dispositivos Universal Serial Bus (USB). As biometrias físicas (impressão digital, íris, retina ou geometria da mão por exemplo), para além de serem um pouco intrusivas, requerem a aquisição de equipamento caro.

Uma possível solução para os problemas enumerados são as biometrias comportamentais. A forma como nos comportamos e agimos num computador pode ser usada como informação biométrica. Esta informação pode ser utilizada *à posteriori*, geralmente complementada com outros métodos, para identificar, inequivocamente, (ou pelo menos com um elevado grau de confiança) um indivíduo. A informação recolhida pode variar desde o tipo de escrita no teclado, habilidade com o rato, hábitos, cliques, número de páginas abertas, origem do acesso, etc. A informação é posteriormente processada por algoritmos comportamentais para criar um perfil de utilizador e usada no processo de autenticação, e de forma contínua durante o tempo de sessão.

2. BIOMETRIAS E DETETORES DE INTRUSÕES BASEADOS EM COMPORTAMENTO

Os sistemas de deteção de intrusões (IDS) são, habitualmente, de dois tipos: baseados em conhecimento ou em comportamento. Neste trabalho debruçamo-nos na deteção de intrusões baseada em comportamento. Convém notar que, atualmente existem poucas soluções que implementem esta abordagem, mesmo que em [2] se reconheça que é um requisito importante para um sistema de deteção de intrusões.

Os sistemas de deteção de intrusões baseados em comportamento assumem que uma intrusão pode ser detetada através da observação de um desvio de um comportamento normal ou esperado do sistema ou dos utilizadores. O modelo de comportamento normal é extraído a partir da informação recolhida, durante a fase de aprendizagem, através de vários meios. O detetor de intrusões compara a atividade atual com o modelo e se um desvio for observado, gera um alarme. Isto significa que qualquer atividade que não corresponda a um comportamento previamente aprendido é considerada intrusiva.

Este tipo de abordagem tem a vantagem de conseguir detetar tentativas de intrusão e vulnerabilidades novas e imprevistas, sendo, ao mesmo tempo, menos dependente dos mecanismos específicos dos siste-

>>>

mas operativos, e assim ajudar a detetar ataques de "abuso de privilégios" que não envolvem nenhuma vulnerabilidade específica. Em suma, esta abordagem mitiga os problemas enunciados na expressão: "Tudo o que não foi visto anteriormente é uma ameaça" [3].

A elevada taxa de falsos alarmes é geralmente citada [3] como a grande desvantagem das técnicas baseadas em comportamento porque todo o âmbito do comportamento de um sistema de informação pode não ser coberto na fase de aprendizagem. Além disso, os comportamentos variam com o tempo causando a necessidade de treino periódico do perfil de comportamento [3].

As biometrias comportamentais têm o potencial de verificar a identidade dos utilizadores com base na sua interação com o computador, maioritariamente com o rato e teclado. Existem várias aplicações de biometrias comportamentais baseadas em interações humano-computador, neste trabalho iremo-nos focar nas seguintes:

- Verificação no início de sessão: Sempre que um utilizador tenta aceder a um serviço, a forma de escrita do seu nome de utilizador e palavra-chave;
- Verificação Contínua: Depois de um início de sessão bem-sucedido, toda a interação do utilizador como computador é monitorada.

Ambas as aplicações fazem uso de técnicas como *keystroke dynamics* (dinâmica da digitação) e *mouse dynamics* (dinâmica do rato), sendo que todas as atividades são continuamente vigiadas de modo a detetar potenciais intrusões.

2.1 *KEYSTROKE DYNAMICS*

Keystroke dynamics [4] é uma biometria baseada na suposição de que pessoas diferentes digitam num teclado de maneiras únicas. Observações de operadores de telégrafo no século XIX revelaram padrões distintos de escrita de mensagens sobre linhas telegráficas, e os seus operadores conseguiam reconhecer-se entre si baseados nos padrões de escrita.

Os algoritmos de *keystroke dynamics* monitorizam a entrada do teclado milhares de vezes por segundo com o objetivo de identificar utilizadores com base no seu habitual padrão de escrita. Este método biométrico, ao contrário da maior parte dos outros, é quase gratuito – o único *hardware* necessário é o teclado. Várias propriedades podem ser extraídas a partir do comportamento de escrita do utilizador, incluindo a latência entre teclas consecutivas, tempo de voo, tempo de pressão, velocidade de digitação e frequência de erros. Para a verificação de textos longos são, normalmente, utilizados *di-grafos, tri-grafos* ou *n-grafos* que ilustram a latência entre duas, três ou n teclas consecutivas, respetivamente.

2.2 *MOUSE DYNAMICS*

Verificar a identidade de um utilizador através das atividades do rato é uma abordagem mais complexa, quando comparada com o mecanismo anteriormente apresentado. O estudo pioneiro e que melhor retrata esta técnica é [5] e define quatro tipos de ações diferentes: movimento, arrastar e largar, apontar e clicar e silêncio. A partir destas ações é possível definir propriedades como interpolações entre a velocidade de movimento e a distância percorrida, que estima a velocidade média em que um utilizador executa certa distância. Adicionalmente, também se podem construir múltiplos histogramas que representam diferentes estratégias de trabalho como a velocidade de movimento média por direção ou a percentagem de ocorrência de cada ação.

3. REFORÇO DOS ATUAIS SISTEMAS DE AUTENTICAÇÃO

A presente secção apresenta um sistema de reforço ao método de autenticação através da introdução de nome de utilizador e palavra-chave. A adição ao método tradicional é a utilização de algoritmos comportamentais para verificar se a forma como o utilizador interagiu com o teclado, durante a introdução do nome de utilizador e palavra-chave, se enquadra no perfil criado anteriormente (através de um processo de aprendizagem) – o nome de utilizador é a chave primária para o perfil.

Para construir um padrão de comportamento de escrita no teclado e o correspondente perfil de interação, todos os eventos produzidos pelo teclado são monitorados de modo. A partir daí, todos os dados recolhidos são avaliados pelas seguintes métricas:

- Tempo de pressão: tempo total que uma tecla é premida;
- Latência: tempo entre carregar numa tecla e largar a seguinte;
- Velocidade: tempo total de escrita do nome de utilizador e palavra-chave;
- Transição: uso da tecla *TAB* ou do rato para trocar entre as caixas de texto;
- Maiúsculas: uso da tecla *SHIFT* ou *CAPS LOCK* nos carateres maiúsculos.

A cada momento de autenticação, o módulo de verificação compara as métricas recolhidas com o

>>>

perfil e responde de forma positiva se a classificação estiver dentro do perfil do utilizador.

A Figura 1 apresenta o fluxograma do processo de autenticação. Apenas os elementos dentro da caixa cinzenta fazem parte do sistema desenvolvido. As restantes funções são externas ao sistema e farão parte do sistema de autenticação – por exemplo IAM.

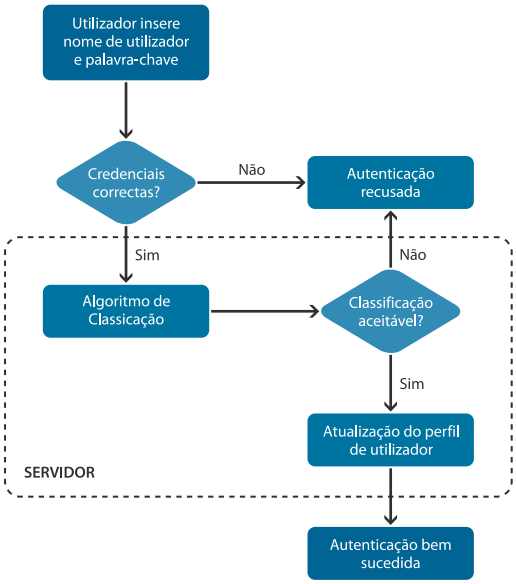


Figura 1. Algoritmo de verificação

O perfil de utilizador é construído nas primeiras 15 sessões e ajustado a cada autenticação, através de utilização de um algoritmo adaptativo [6]. A utilização deste algoritmo permite que o sistema se adapte conforme as mudanças de comportamento do utilizador. Sempre que existir uma autenticação correta esta é adicionada ao perfil de comportamento – as métricas mais antigas são removidas e a nova introduzida. Este deslizamento temporal permite uma adaptação mais coerente com alterações temporais de comportamento, garantindo que o perfil de utilizador está ao corrente das mudanças comportamentais do utilizador da conta (i.e. se um acidente acontece e altera o comportamento do utilizador, o algoritmo irá adaptar-se à mudança num curto período de tempo).

4. AUTENTICAÇÃO CONTÍNUA

O senso comum leva-nos a considerar que as intrusões só acontecem por vias externas, mas grande parte das ameaças corporativas são baseadas em utilizadores internos e próximos à vítima [7]. A falta de verificação contínua da identidade do utilizador é uma grave vulnerabilidade, já que cerca de 70% dos ataques cibernéticos surgem de intrusos internos [7]. Por isso, é crucial monitorar e agir já depois da fase de autenticação para mitigar alguns destes problemas. O sistema de autenticação contínua foi desenvolvido tendo em consideração estes problemas, proporcionando um método para detetar padrões de comportamento anormais. Através da monitorização constante e contínua do rato e teclado é obtida informação que pode ser usada para detetar uma possível intrusão.

A Figura 2 apresenta uma visão geral do sistema. Este sistema é bastante mais complexo que o apresentado na secção anterior. O facto é que a validação contínua traz dois problemas: i) a capacidade de processamento e ii) o tempo de resposta, isto é, o sistema tem de ser capaz de capturar todas as interações do utilizador e processá-las.

Os dados a partir do teclado e do rato podem ser combinados, levando à obtenção de resultados mais precisos. No entanto, os métodos utilizados para capturar dados de cada um dos dispositivos são muito diferentes. As secções a seguir apresentam as metodologias de monitorização utilizadas.

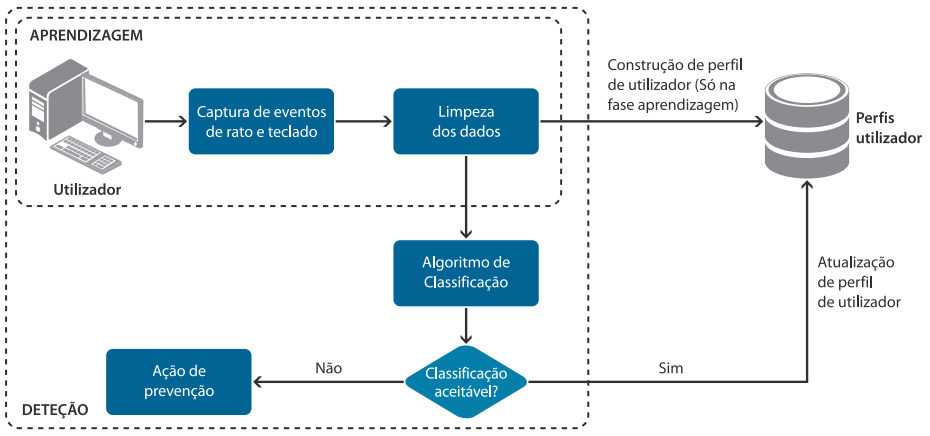


Figura 2. Sistema de autenticação contínua

4.1 MONITORIZAÇÃO DO TECLADO

A monitorização do teclado é efetuada da mesma forma da do sistema de autenticação apresentado anteriormente. O sistema recolhe os dados, filtra os eventos produzidos por teclas especiais (que não têm significado na produção de texto) e depois processa esses mesmos dados com base num algoritmo que é capaz de classificar texto contínuo e livre. As métricas usadas são:

- Tempo de pressão: herdado do sistema anterior;
- Latência de uma palavra: são medidas todas as latências das combinações das n letras da palavra (n-grafo).

A tabela 1 apresenta um exemplo do cálculo da latência para a palavra "TESE". De modo a calcular os diversos grafos da palavra são capturados os *timestamp* dos eventos de pressão e libertação de cada tecla, e a partir daí são calculadas as latências (em milissegundos) das letras da palavra.

KEYSTROKES			DÍGRAFO		TRI-GRAFO		TETRA-GRAFO	
Tecla	Premida	Largada	Grafo	Latência	Grafo	Latência	Grafo	Latência
T	798340	798409	T + E	187	T + E + S	358	T + E + S + E	510
E	798403	798527	E + S	295	E + S + E	447		
S	798605	798698	S + E	245				
E	708746	708850						

Tabela 1. Exemplo de n-grafo

A fase de aprendizagem do algoritmo corresponde às quinze primeiras sessões do utilizador no sistema. Depois é aplicado o algoritmo adaptativo apresentado anteriormente. Cada sessão é constituída por 250 caracteres.

De modo a tornar o sistema mais robusto e completo foi também implementado um módulo de monitorização de rato, sendo apresentado de seguida.

4.2 MONITORIZAÇÃO DO RATO

A monitorização do rato é efetuada através do registo de todos os eventos de movimento e clique. Os dados são recolhidos a cada 100 milissegundos e constituídos pelas seguintes ações:

- Silêncio: ausência de movimento;
- Movimento: movimento do rato:
- Apontar e clicar: movimento seguido de clique ou duplo-clique;
- Arrastar e largar: movimento durante um clique.

Num sistema *multi-task*, é expectável que o comportamento do utilizador seja diferente de aplicação para aplicação. O movimento do rato está também relacionado com o tamanho do ecrã, já que este influencia o cálculo da distância. Assim, devido ao fluxo de trabalho das aplicações em que o utilizador opera, são obtidos diferentes padrões de uso. Para ultrapassar este problema, cada aplicação tem o seu próprio perfil de comportamento. Este método proporciona um resultado mais preciso do que uma captura geral. Por exemplo, escrever um documento tem um padrão de interação com o rato que é muito diferente de desenhar uma imagem.

De modo a reduzir a enorme quantidade de dados recolhidos, foi usado a fórmula de amplitude interquartil [8] que deteta e elimina os valores atípicos (*outliers*) presentes nos dados. Estes *outliers* podem ser erros de leitura, erros de interpretação ou comportamentos dispersos que não traduzem qualquer significado específico. Após a utilização desta fórmula verificou-se uma eliminação de cerca de 25% dos dados o que aumentou a eficiência do sistema.

Os próximos pontos demonstram as métricas possíveis de extrair de um perfil de utilizador:

- Velocidade de movimento comparado com a distância;
- Percentagem de movimento por direção;
- Velocidade média por direção;
- Percentagem de tipos de ação;
- Velocidade média por ação.

A Figura 3 mostra os dados recolhidos em cru e a respetiva curva de perfil. Os pontos a azul correspondem aos dados brutos. Como estes pontos são dispersos e de difícil interpretação, estes são aproximados através de um curva polinomial (curva amarela) utilizando um algoritmo de regressão polinomial de grau 15.

>>>

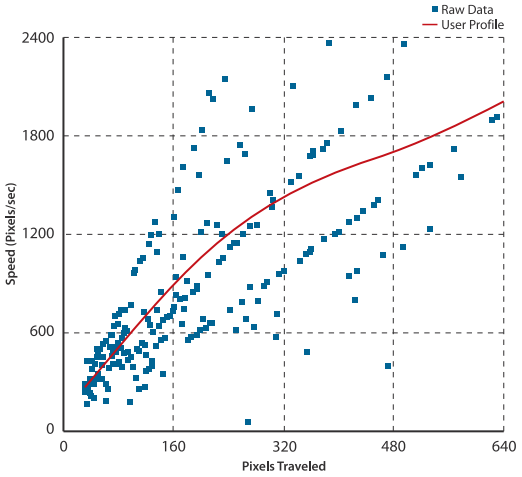


Figura 3. Velocidade por distância de movimento

A curva resultante é muito precisa, contendo todas as características e ações comuns do utilizador. Como resultado, uma intrusão pode ser detetada quando a curva de comportamento atual for diferente da curva de comportamento normal, como se pode ver na figura 4.

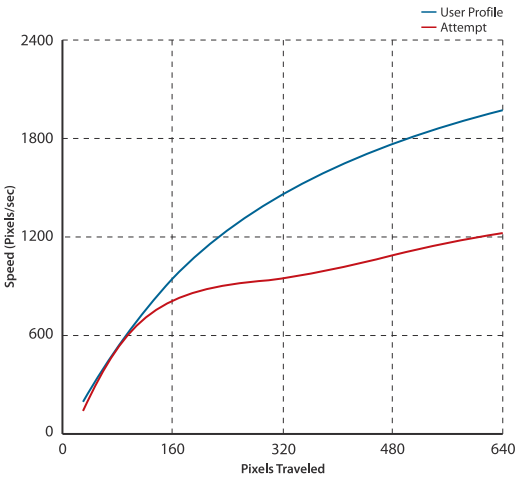


Figura 4. Sessões de utilizadores distintos

Além disso, na figura 4 a linha azul corresponde ao perfil de comportamento normal do utilizador e a linha amarela uma tentativa de intrusão. A distinção é muito clara e as curvas diferem muito em termos de velocidade e distância. Em resultado, a deteção é feita e é gerado um alerta.

Na figura 5 é mostrado a precisão do sistema e a geração de uma curva de utilização atual, e a curva de perfil do utilizador. Em particular, a curva de perfil é construída após 15 sessões de interação. Embora haja algumas diferenças mínimas entre as curvas, estas apresentam um modelo muito semelhante pelo que o sistema comporta-se de forma normal.

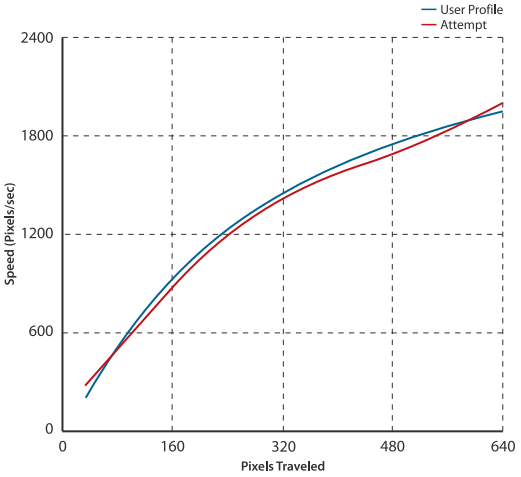


Figura 5. Sessões do mesmo utilizador

5. RESULTADOS

Os casos de estudo foram realizados num ambiente controlado, onde os utilizadores tinham em conta que todos as suas interações iriam ser exaustivamente analisadas de maneira a verificar a eficiência dos algoritmos de classificação e dos módulos de recolha de dados.

Para isso, foram desenvolvidos dois casos de estudo, um para o sistema de autenticação estática e outro para o sistema de autenticação contínua. O ambiente escolhido foi um laboratório informático.

5.1 ALGORITMOS DE CLASSIFICAÇÃO

A escolha do algoritmo de classificação adequado é um dos fatores mais importantes num sistema de deteção de intrusões. Este algoritmo é responsável por classificar as sessões em relação ao perfil do utilizador. A classificação resultante da execução deste algoritmo determina se o autor da sessão é ou não fidedigno. Em [9] é efetuado um estudo comparativo utilizando 14 algoritmos distintos. Estes algoritmos são utilizados num problema de *keystroke dynamics* semelhante ao desenvolvido neste trabalho.

Para o caso de sistemas biométricos existem dois tipos de erro que podem ser usados para medir a sua eficácia. Esses tipos de erro são os seguintes:

- Erro Tipo I: percentagem de utilizadores fidedignos que são rejeitados pelo sistema;
- Erro Tipo II: percentagem de impostores que foram aceites pelo sistema.

Quanto menor a percentagem de ambos os erros, mais forte e eficaz é o sistema biométrico.

Devido à complexidade do algoritmo e aos seus resultados, neste trabalho foram adotados os algoritmos *Outlier Count* e *Nearest Neighbor* (Mahalanobis) [9].

5.2 CASO DE ESTUDO – REFORÇO DE AUTENTICAÇÃO

O primeiro caso de estudo diz respeito ao sistema de autenticação estática. Foi criada uma página Web com um formulário de início de sessão simples e algumas contas de utilizadores para teste. O sistema foi testado por 10 utilizadores em que cada um escolheu um nome e palavra-chave livremente e realizou 30 tentativas de iniciar sessão no sistema, usando as suas credenciais. No fim foi calculada a taxa de falsos negativos (erro tipo I), ou seja, a quantidade de tentativas que um utilizador válido não conseguiu aceder à sua conta.

Depois, a totalidade das credenciais foram partilhadas e todos os utilizadores tentaram iniciar sessão 2 vezes para cada uma das contas que não eram suas. Com este teste calculou-se a percentagem de falsos positivos (erros tipo II) ou seja a quantidade de vezes em que uma conta foi acedida por um intruso.

No total foram realizados 500 tentativas de início de sessão (300 legítimas e 200 de tentativa de intrusão) por 10 pessoas num período de uma semana.

Algoritmo	Erro Tipo I para Tipo II = 0%	Taxa de Erro Igual	Erro Tipo II para Tipo I = 0%
Outlier Count	19%	4,9%	32,4%
Mahalanobis	16,6%	6,5%	30,6%
Combinado	**10%**	**4,2%**	**27%**

Tabela 2. Resultados do sistema de autenticação estática

Na tabela 2 é apresentada a comparação entre os algoritmos para este problema. Na primeira coluna é apresentada a percentagem de erros de tipo I quando o erro de tipo II é 0%, cenário onde nenhum intruso consegue quebrar o sistema. Analogamente, a última coluna apresenta a percentagem de erros de tipo II quando o erro de tipo I é 0%, significando que o utilizador fidedigno nunca é rejeitado pelo sistema. A coluna do meio apresenta a taxa de erro igual, que é a taxa de erro escolhida.

Analisando a tabela, o algoritmo combinado apresenta os melhores resultados. Como os algoritmos *Outlier Count* e *Mahalanobis* são melhores em diferentes casos, o algoritmo combinado (que surge na fusão dos dois anteriores) apresenta um resultado global positivo para este projeto, e por conseguinte, foi o algoritmo escolhido na implementação final.

Além disso, ao longo dos testes, os utilizadores foram autorizados a escolher o seu nome de usuário e palavra-chave livremente, com tamanhos entre os 5 e 10 caracteres, o que resultou numa grande variação nos resultados. A maior parte dos erros de tipo II foram despoletados por utilizadores com palavras-chave curtas, sendo que os melhores resultados foram obtidos por utilizadores com credenciais compostas por 8 ou mais carateres.

5.3 CASO DE ESTUDO – SISTEMA DE AUTENTICAÇÃO CONTÍNUA

Neste cenário de utilização foi instalado nativamente o sistema de identificação contínua em 6 computadores, sendo estes computadores operados só por uma pessoa. O sistema foi executado durante uma semana com a pessoa fidedigna para se calcular a taxa de erro de tipo I.

Depois cada pessoa usou a conta de outros durante um dia para assim conseguirmos obter a taxa de erro de tipo II.

No total foram realizadas 525 sessões legítimas (390 de teclado e 135 de rato) e 205 sessões de tentativa de intrusão (135 de teclado e 70 de rato). O número de sessões de teclado é significativamente maior em relação às de rato já que o tempo necessário, em utilização contínua, para a construção de uma sessão de teclado ronda os 2 minutos contra os 10 minutos de uma sessão de rato. Quanto mais baixa a taxa de ambos os erros, melhor a eficiência do sistema.

Para uma melhor comparação dos resultados obtidos, a tabela 3 apresenta os resultados para cada um dos algoritmos, seguindo a forma de apresentação da tabela 2.

Algoritmo	Dispositivo	Erro Tipo I para Tipo II = 0%	Taxa de Erro Igual	Erro Tipo II para Tipo I = 0%
Outlier Count	Teclado Rato	53% 48,7%	**4,2%** 20,3%	25% 25,3%
Mahalanobis	Teclado Rato	55,7% 31%	12,6% 12,5%	25% 20%
Combinado	Teclado Rato	**44%** **12,4%**	9% **5,9%**	**24,6%** **11%**

Tabela 3. Resultados do sistema de autenticação contínua

Mais uma vez, o algoritmo combinado é o que apresenta melhores resultados, sendo superior em cinco dos seis cálculos. Esta abordagem é realmente a melhor num cenário normal e para a maior parte dos utilizadores.

Todas estas iterações são muito positivas, devido à complexidade e o volume de cálculos que este sistema exige.

6. CONCLUSÕES E TRABALHO FUTURO

O uso de biometrias comportamentais apresenta claras vantagens como um método para reforço dos tradicionais métodos de autenticação. A sua transparência, aceitabilidade e a não implicação de mudanças de rotinas, fazem destas técnicas uma escolha inteligente para a construção de sistemas de deteção de intrusões. Neste trabalho foi apresentado um sistema completo que usa duas abordagens distintas para a autenticação contínua do utilizador, mouse dynamics e *keystroke dynamics*.

Os resultados finais, bastante satisfatórios, mostram que o sistema de autenticação apresenta resultados a par com os projetos concorrentes [1] [4] [5] [7], e o sistema de autenticação contínua apresenta resultados muito promissores. Há que evidenciar, que é necessário um compromisso entre a permissividade (ou não) do algoritmo e a complexidade do texto – texto (i.e. *password*) mais complexo, poderá permitir uma maior permissividade do algoritmo de validação, pelo contrário texto (i.e. *password*) menos complexo obrigará a uma menor permissividade do algoritmo de validação.

A criação de perfis e a utilização de algoritmos comportamentais estão geralmente associados a problemas relacionados com o facto do ser humano não ser (muitas vezes) previsível, ou o ambiente que o rodeia poder, em certas ocasiões, alterar-lhe o comportamento. Nos cenários apresentados, o facto de o utilizador estar com elevados níveis de *stress* ou num ambiente diferente (desfavorável – por exemplo uma viagem de comboio) poderá limitar a utilização das técnicas e algoritmos apresentados neste artigo. É por isso importante reconhecer que estas técnicas deverão ser sempre usadas como complemento de outras técnicas de autenticação (por exemplo, autenticação baseada em contexto – geografia/rede de acesso, autenticação física, …).

O sistema desenvolvido foi integrado com o produto IAM, servindo para que se acione um outro método de autenticação (*one-time-token via Short Message System (SMS)*) no caso do algoritmo de validação retornar uma resposta negativa. A filosofia foi a de obrigar à utilização de outro fator de autenticação e não à recusa de acesso. Outras combinações são obviamente possíveis!

Sendo este um fator importante para a melhoria da eficiência dos sistemas, a implementação de algoritmos de deteção de *stress* para a minimização de erros relacionados com variações de comportamento está considerado como trabalho futuro.

No sistema de autenticação contínua é fundamental o utilizador saber que está a ser monitorado de uma forma contínua. Todos os caracteres e todo o movimento do rato são guardados e analisados, o que levanta sérias questões de privacidade. A utilização deste tipo de sistemas de forma generalizada não é de todo possível, sendo expectável a sua utilização em determinados cenários – por exemplo, infraestruturas criticas militares, de segurança ou outras. De qualquer forma, os autores têm vindo a trabalhar no sentido de tentar garantir que toda a informação recolhida é cifrada com chave, conhecida apenas pelo utilizador, mantendo todas as funcionalidades e benefícios apresentados ao longo deste artigo. É um processo complexo de criptografia em que o algoritmo de validação é capaz de executar os cálculos, sem decifrar a informação necessária para os cálculos.

Finalmente, é de referir que o trabalho realizado é adaptável a teclados virtuais e ecrãs multitoque.

REFERÊNCIAS

[1] L. O'Gorman, "Comparing Passwords, Tokens, and Biometrics for User Authentication," *Proceedings of the IEEE*, Vol. 91, No. 12, pp. 2019-2040, 2003.

[2] D. E. Denning, "An Intrusion-Detection Model," em *IEEE Transactions on Software Engineering - Special issue on computer security and privacy*, IEEE Press Piscataway, 1987, pp. 222-232.

[3] H. Debar, "Intrusion Detection FAQ: What is behavior-based intrusion detection?," 19 Maio 2010. [Online]. Available: http://www.sans.org/security-resources/idfaq/behavior_based.php.

[4] J. Ilonen, "Keystroke dynamics," em *Advanced Topics in Information Processing-Lecture*, 2003.

[5] A. A. E. Ahmed e I. Traore, "A New Biometric Technology Based on Mouse Dynamics," em *IEEE Transactions on Dependable and Secure Computing*, 2007, pp. 165-179.

[6] I. Koychev e I. Schwab, "Adaptation to Drifting User's Interests," em *In Proceedings of ECML2000 Workshop: Machine Learning in New Information Age*, 2000, pp. 39-46.

[7] J. Ferreira, H. Santos e B. Patrão, "Intrusion detection through keystroke dynamics," em *The Proceedings of the 10th European Conference on Information Warfare and Security*, Tallin, 2011, pp. 81-90.

[8] NIST/SEMATECH, e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 2012.

[9] K. S. Killourhy e R. A. Maxion, "Comparing Anomaly-Detection Algorithms for Keystroke Dynamics," In *Proceedings of DSN*, pp. 125-134, 2009.

CVS DOS AUTORES

Henrique Fontão Martins, licenciado e mestrado em Engenharia Informática na Universidade do Minho em 2011 e 2013 respetivamente. Estagiário da PT Inovação desde fevereiro de 2013, onde desenvolveu a sua tese de mestrado que originou este artigo. Atualmente está ligado a projetos de autenticação, autorização e controlo de acessos.

Ricardo Azevedo, MSC em Internet Computing pelo Queen Mary College (Universidade de Londres), em 2006, e Licenciado em Engenharia de Computadores e Telemática pela Universidade de Aveiro, em 2004. Foi bolseiro de investigação do IT desde 2004 a Fevereiro de 2006, com trabalho desenvolvido na área de QoS em redes heterogéneas de 4ª geração. Em Fevereiro de 2006 iniciou a sua actividade profissional na PT Inovação. Tem participado em diversos projectos de I&D no âmbito do IST e Eurescom, nas áreas de QoS e Network Management, Mobilidade, Segurança, Privacidade e Gestão de Identidades. Atualmente é Team Leader do grupo de Gestão de Identidades e Privacidade.

09 Privacidade, a *M2M killer-issue*?


JOÃO GONÇALVES FILIPE PINTO

PALAVRAS CHAVE
Privacidade, M2M, *Ambient Intelligence*

A tecnologia aproxima-se a passos largos de ser capaz de realizar a visão de *Ambient Intelligence* [1] que serviu de orientação à investigação europeia na última década. Cada utilizador interagirá diariamente com centenas de dispositivos que terão ligação de rede. Muitas destas interações não serão explícitas – imaginemos a detecção da passagem de uma pessoa num corredor através da utilização de um detector de movimentos – pois muitos destes dispositivos estarão equipados com sensores que, sem intervenção humana, recolhem dados do ambiente que os rodeia. Esta informação será então usada na provisão, adaptação e personalização de serviços. As comunicações máquina-a-máquina (M2M) aparecem como um dos *enablers* deste cenário, sendo apontado como o negócio futuro mais relevante para as empresas de telecomunicações. No entanto esta grande quantidade de informação em circulação, conjugada com custos de armazenamento de dados em queda, levanta questões de privacidade que vão desde problemas pontuais de intimidade até à vigilância Orweliana [2]. Estes medos de invasão de privacidade são confirmados pelo comportamento atual das empresas americanas de crédito, seguros e publicidade *web*, que transacionam livremente os dados pessoais dos utilizadores, a maioria das vezes sem o seu conhecimento [3].

A privacidade é uma questão relevante nos dias que correm, e será cada vez mais à medida que o digital se funde com o físico. É também um conceito frequentemente mal traduzido em requisitos e quase sempre mal implementado [4]. Por essa razão este artigo começará por analisar as questões socioculturais e legais da privacidade, nomeadamente a demarcada diferença entre o entendimento Europeu e Americano sobre o assunto. De seguida ir-se-á focar na implementação de cenários de *Ambient Intelligence* através de uma infraestrutura M2M, e levantará as questões de privacidade mais relevantes. As tecnologias existentes que permitem proteger a privacidade do utilizador são apresentadas e finalmente são conjeturadas soluções de privacidade que possam ser implementadas nessa infraestrutura M2M.

1. INTRODUÇÃO

Os operadores de telecomunicações não querem perder as oportunidades geradas pela massificação das comunicações máquina-a-máquina (M2M). Todos os estudos efetuados apontam para um crescimento exponencial deste novo mercado assente em dispositivos que comunicam sem intervenção humana. A informação recolhida por uma variedade de sensores aliada à possibilidade de actuação no ambiente real abre portas a uma infinidade de novos serviços que englobam os mais diversos sectores de actividade, de onde se destacam as áreas da telemetria, controlo remoto, saúde, transporte e, por fim, segurança.

As comunicações máquina-a-máquina potenciam um mercado de vários milhares de milhões de euros onde os operadores de telecomunicações querem estar de corpo presente. Para isso, terão de se tornar facilitadores dos negócios de todos os intervenientes no ecossistema M2M. Os operadores de telecomunicações poderão assegurar as comunicações entre dispositivos, facilitar a mediação de dados M2M a aplicações de terceiros e ainda providenciar mais e melhores serviços.

No entanto, associados aos novos paradigmas de comunicação surgem sempre novos desafios que não podem ser desprezados. As questões de privacidade têm estado, ultimamente, na ordem do dia, vindo as comunicações M2M agudizar este problema. A informação gerada pelas máquinas pode ser particularmente intrusiva na intimidade das pessoas, principalmente quando cruzada com outros tipos de informação. Será a privacidade, o *M2M killer-issue*?

Nos últimos anos são muitas as notícias de problemas de privacidade relacionados com as tecnologias de informação e comunicação (TIC). Em 2006 a AOL disponibilizou temporariamente, para fins de investigação, uma base de dados com os termos de pesquisa de milhões de utilizadores. Embora os utilizadores não estivessem diretamente identificados, os termos de pesquisa libertados que referenciavam locais, serviços e pessoas permitiram que inúmeros desses utilizadores fossem identificados, e permitiram associar-lhes termos de pesquisa sensíveis [5]. Em 2011 a Sony anunciou que tinha sofrido um ataque aos seus servidores de jogo online, usados pela consola Playstation, e que os dados pessoais de milhões de utilizadores, incluindo nomes, moradas, endereços de correio eletrónico e números de cartão de crédito, poderiam ter sido comprometidos [6]. Já neste ano de 2013 os conhecidos PRISM leaks tornaram públicas as obrigações legais que várias empresas privadas norte-americanas têm relativas à transmissão de dados para agências de informação, espionagem e segurança. A operadora americana Verizon tem de fornecer diariamente informações sobre as chamadas efetuadas [7], e os gigantes da Internet têm de fornecer informação a pedido sobre pessoas ou conceitos [8].

Já há mais de 100 anos que o progresso tecnológico obriga à discussão sobre o direito à privacidade. Em 1888 a Kodak lançou a primeira máquina fotográfica que permitia tirar fotografias com o simples carregar de um botão. Dois anos depois surgia a discussão sobre a necessidade de haver leis sobre o direito à privacidade, que na altura era visto na prática como impedir que pessoas fossem fotografadas contra a sua vontade. O período com maior relevância para o desenvolvimento de regras de

>>>

privacidade foi a década de 70. O surgimento dos *mainframes*, onde os registos com informações pessoais começaram a ser guardados eletronicamente, levantou questões sobre as possibilidades de abuso da informação recolhida. O debate gerou a criação de um conjunto de normas a seguir ao lidar com informação pessoal de outrem, denominadas *Fair Information Practices* (FIPs).

Figura 1. Máquina Fotográfica Kodac de 1888

Parece estar próximo outro momento de definição histórico relativamente às regras de privacidade aceites pela sociedade. As questões mais relevantes estão a ser analisadas por políticos e magistrados, com especial relevância para os trabalhos desenvolvidos na União Europeia. No entanto resta saber se estes esforços revelar-se-ão suficientes para o futuro próximo, mediando convenientemente os interesses económicos e sociais num cenário de *Ambient Intelligence*.

2. ANÁLISE DO ESTADO ATUAL

PAPÉIS DOS OPERADORES DE TELECOMUNICAÇÕES NO M2M

As comunicações máquina-a-máquina darão origem a um novo mundo de negócios inovadores, onde a informação e o controlo providenciados por dispositivos ligados serão a base para uma nova gama de serviços mais inteligentes. Os operadores de telecomunicações têm aqui uma oportunidade de ouro para alargar a sua gama de ofertas, potenciado a criação de mais receitas. Eles terão de desempenhar papéis muito específicos de forma a tornarem-se nos intervenientes principais neste mercado gigantesco, em particular:

- *Gestão de Conectividade* – Os operadores de telecomunicações terão de garantir a transmissão de informações M2M de forma transparente entre as extremidades, garantindo a qualidade de serviço acordada e as necessárias características de segurança do canal de comunicação. Nestes casos, os operadores são pouco mais do que *bit pipes*, transmitindo apenas a informação sobre as suas redes, potenciando a utilização dos seus sistemas, acompanhando o aumento da quantidade de dispositivos conectados.

- *Mediação de Dados* – A enorme quantidade de novos dispositivos acabará por gerar grandes volumes de dados que necessitam de armazenamento e processamento. Os operadores de telecomunicações farão assim a ponte entre aplicações e dispositivos heterogéneos, permitindo a criação rápida de novos serviços, assegurando a normalização das interfaces. Embora a informação seja ainda opaca para o operador, este terá de garantir a segurança dos dados, protegendo a privacidade dos utilizadores.

- *Fornecedor de Serviços* – O operador pode ainda criar os seus próprios serviços, enriquecendo-os com a informação disponibilizada pelos dispositivos, criando verdadeiras ofertas de produtos inteligentes. Pode ainda trabalhar os dados M2M fornecendo informações de valor acrescentado a entidades interessadas, que podem agora melhorar os seus processos com novos recursos e funcionalidades.

A Figura 2 apresenta os papéis principais que um operador de telecomunicações poderá ter num ecossistema M2M.

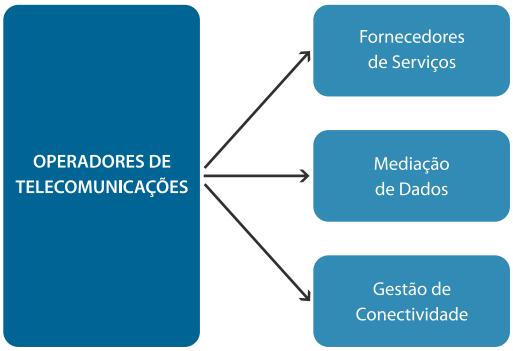


Figura 2. Papéis dos operadores de telecomunicações num ecossistema M2M

PRIVACIDADE, ECONOMIA DIGITAL E VIGILÂNCIA

"No mundo de hoje a informação pessoal é a moeda da economia digital." Quem o disse foi a Vice-Presidente da Comissão Europeia, Viviane Reding, por ocasião do anúncio da Reforma Europeia de Proteção de Dados [10]. A informação sobre a utilização da

>>>

web é usada para fornecer publicidade direcionada, representando um mercado de milhares de milhões de dólares que representa grande parte das receitas das grandes empresas *web* [11]. Mas a informação pessoal não é apenas rentabilizada indiretamente mas também diretamente: há empresas, denominadas *data brokers*, cujo negócio é vender informação de consumo real para análises de mercado e de campanhas de marketing. Uma dessas empresas, a Acxiom, tem um volume de negócios acima dos mil milhões de dólares [12]. A regulação deste mercado é feita de uma forma relativamente permissiva nos Estados Unidos. A *Federal Trade Commission* exige que as empresas cumpram as FIPs, mas a fiscalização é praticamente inexistente, agindo na maioria das vezes reactivamente ao surgir publicamente uma questão de violação de privacidade [13, p. 72].

Na Europa, ao contrário dos Estados Unidos, as leis de privacidade afetam de forma semelhante as instituições públicas e privadas, sendo mais restritivas no velho continente. Talvez por esse motivo as grandes empresas que operam em mercados de informação pessoal sejam quase todas norte-americanas. A reforma Europeia de proteção de dados, iniciada pela Comissão Europeia em 2013, tem como objetivo aumentar a competitividade das empresas europeias na economia digital. Ter um conjunto claro de regras de privacidade para toda União e uma população que confia que os seus dados estão adequadamente protegidos, poderão ser os catalisadores necessários para o surgimento de grandes empresas europeias no sector [10].

A acumulação de dados pessoais representa riscos acrescidos para os *data subjects* (indivíduos aos quais a informação se refere). Quanto maior acumulação, mais aliciante se torna ter acesso às bases de dados dessas empresas, tanto para atores que funcionam à margem da lei (*hackers*), como para atores legais que a usam (agências de segurança e espionagem). Mesmo o risco de abusos cometidos, quer institucionalmente pela empresa, quer por colaboradores (à margem das políticas da empresa), sobe à medida que o valor da informação agregada aumenta. Esse valor tem um potencial de crescimento grande, apenas função da adoção e utilização dos sistemas. Os ataques informáticos têm um custo associado, mas quanto maior for a recompensa, mais *hackers* se dedicarão a tentar consegui-la, usando mais recursos para o ataque. Coincidência ou não, o prejuízo para empresas inglesas resultante de ataques informáticos triplicou num ano [14]. Nesse sentido, o crescimento na dimensão da base de dados tem de ser acompanhado por investimentos nos mecanismos de segurança e auditoria das empresas.

Para além de aumentar o risco de ataques, a agregação de informação pessoal pode ser ilegal segundo as leis Portuguesas e Europeias. A Reforma Europeia da Proteção de Dados prevê três tipos de informação pessoal: informação pessoalmente identificável, informação pseudónima e informação anónima. Para cada um destes tipos de informação há diferentes requisitos legais sobre o seu tratamento. Embora o diploma ainda esteja a ser revisto, espera-se que, dependendo da forma como é armazenada, a informação seja legalmente classificada de formas diferentes.

As práticas de vigilância governamental expostas pelos *PRISM leaks* são simultaneamente um risco para as empresas do sector, pelo medo de partilha de informação que pode suscitar nos consumidores, e uma oportunidade para empresas Europeias se diferenciarem das Americanas pelo simples facto de não estarem sujeitas aos ditos mecanismos. Embora seja provável que existam mecanismos semelhantes na Europa, o facto de ser um governo local, para o qual os cidadãos votam, e não um governo de outro país, poderá tranquilizar os consumidores europeus.

3. DESAFIOS E POSSÍVEIS SOLUÇÕES

PRIVACIDADE NO M2M

Num cenário *web*, a informação é normalmente recolhida através da sua inserção manual pelos utilizadores, ou pelas ações que estes tomam no sistema. Numa aplicação de *smartphone*, para além destas formas, também é possível captar informação dos sensores que o utilizador autorizou (muitas vezes sem se aperceber) a aplicação a usar. Num cenário M2M a informação é captada por um grande número de sensores, distribuídos geograficamente, apenas necessitando de interação humana na instalação e em algumas formas de manutenção. Consequentemente, o potencial de recolha de informação aumenta drasticamente, não só em quantidade como em variedade de tipos, pois os sensores não estão limitados aos existentes nos *smartphone*s. No entanto, novos tipos de informação trazem novas ameaças de privacidade. Por exemplo, estudando ameaças de privacidade do cenário de *Smart Metering*, uma equipa de estudo da Comissão Europeia escreveu [15, p. 5]:

"Dos dados de consumo de energia detalhadamente recolhidos pelos *smart meters*, muita informação relativa aos utilizadores pode ser inferida, tal como a utilização de bens ou dispositivos, rotinas diárias, condições de vida, ocupações, estilo de vida e comportamento."

>>>

Para além do potencial de intrusão acrescido, alguns cenários M2M têm uma cadeia de valor longa, com muitos atores envolvidos, e uma cadeia de distribuição e reutilização da informação complexa. Estes cenários apresentam maior risco de problemas de privacidade do que cenários mais simples, pois mais pessoas e entidades têm acesso à informação recolhida, e a relação com o utilizador e os incentivos de proteger a sua privacidade correm o risco de se diluir pela cadeia. Agravando o problema, os mecanismos de acesso e segurança definidos pelas normas só atuam até a informação chegar a uma entidade aplicacional que interage diretamente com a plataforma M2M. A partir desse ponto não há controlo nem conhecimento do fluxo de informação: ela sai do sistema.

Em termos de segurança de sistemas, para além dos riscos de intrusão central, também existem riscos significativos de ataque de interceção na rede. No entanto são limitados em termos de quantidade por ser necessário que o adversário esteja fisicamente próximo do nodo de rede atacado. Este risco é considerado um risco de segurança M2M e é tipicamente abordado pelas normas propostas.

POSSÍVEIS SOLUÇÕES

Também abordados pelas normas são as questões de autenticação dos dispositivos e cifragem dos dados transportados pela plataforma M2M. Estes mecanismos de segurança essenciais, como foi referido, não cobrem os cenários M2M complexos, e não permitem controlar o fluxo da informação ao longo de toda a cadeia de valor. Para tratar deste problema há algumas tecnologias na área da Gestão de Identidades que podem interessar. Neste campo pode-se adotar uma estratégia de agregação e interoperabilidade entre vários domínios, implementando uma federação, ou pode-se impor uma camada de identidade controlada pela plataforma M2M. Para mais informação sobre estas tecnologias podem-se consultar artigos de edições anteriores da revista Saber&Fazer [16][17][18].

A agregação de informação aumenta os riscos de segurança e cria questões legais para as empresas que a praticam. Nesse sentido pode ser interessante considerar as técnicas existentes para a privacidade no contexto de bases de dados. Já foi demonstrado que simplesmente apagar identificadores, como nomes, e-mails e moradas, não é suficiente para proteger os dados dos utilizadores [19]. No entanto existe uma grande quantidade de técnicas, algumas baseadas em estatística [21], outras relacionadas diretamente com o *data mining* [20], para minimizar os riscos associados à agregação de dados.

Frequentemente as questões de privacidade surgem, não motivadas por questões tecnológicas, mas por questões de negócio. Por um lado, os incentivos do modelo de negócio podem levar a esses problemas, tal como acontece frequentemente com os serviços *internet*. Por outro, as parcerias que o M2M sugere podem confundir o utilizador relativamente aos fluxos de informação no sistema. Os erros de desenho de sistemas e soluções mais comuns que acabam por criar problemas de privacidade são cinco [4]:

- Esconder potenciais fluxos de informação: no caso de a informação ser utilizada para outros fins que não o principal, isso deve ser claro para o utilizador;
- Dificultar a perceção do fluxo normal de informação: quando a informação é partilhada por parte do utilizador deve ser muito claro quem vai imediatamente ter acesso a esta;
- Configuração em vez de ação: a privacidade deve surgir através do uso normal do sistema, não através de uma configuração estática;
- Inexistência de um controlo global: deve haver sempre uma maneira óbvia de interromper globalmente a publicação de informação;
- Ignorar a prática comum: deve haver o cuidado de tentar transferir as práticas sociais de privacidade existentes para o desenho do sistema.

Ter atenção para não cometer estes erros e apresentar ao utilizador final uma visão coerente e clara do fluxo de informação são questões essenciais para uma receção pública positiva de serviços M2M.

4. IMPORTÂNCIA PARA OS NEGÓCIOS DO GRUPO PT

Vários estudos indicam que o M2M dará origem nos próximos anos a um novo mercado de milhares de milhões de euros. Independentemente da exatidão destes valores, parece certo que o volume de negócios associados à comunicação entre máquinas não será desprezável. O operador de telecomunicações deve tornar-se no verdadeiro facilitador de novos negócios independentemente do sector de atividade, tornando-se assim no principal *stakeholder* no ecossistema M2M. Assim, o operador pode promover as suas redes de telecomunicações, fixas ou móveis, assegurando um aumento de tráfego nos seus sistemas. Deve ainda garantir a mediação da informação de uma forma segura, permitindo a aplicações o acesso transparente a dispositivos heterogéneos. Por fim, deverá criar os seus próprios serviços, talhados à necessidade dos mercados, oferecendo soluções inovadoras a futuros clientes.

>>>

No entanto estes novos serviços podem ser mal recebidos pelos utilizadores finais, caso haja dúvidas quanto aos riscos de privacidade. As empresas de serviços *Web* tipicamente não exigem pagamento por parte do utilizador, monetizando o uso que as pessoas fazem do serviço através da publicidade e informação recolhida. Para estas, as questões de privacidade sucedem-se mas os utilizadores não mudam significativamente as suas opções. No entanto, supõe-se que para os operadores de telecomunicações o comportamento dos consumidores será bastante diferente. Será difícil de justificar fugas de informação e problemas de privacidade num serviço pago. Além disso, os operadores têm uma relação com o cliente mais próxima, e transmitem tipicamente uma imagem de confiança e qualidade de serviço. A acrescer a tudo isto, devido ao seu modelo de negócio, um problema de relações públicas pode ter repercussões diretas na faturação.

5. CONCLUSÕES

O M2M surge como um negócio de futuro para os operadores de telecomunicações, no entanto a sua adoção traz riscos significativos de ferir a imagem de confiança transmitida durante anos aos utilizadores. É necessário fazer uma análise cuidada tanto ao nível de desenho do negócio como do sistema, e também do uso das tecnologias mais apropriadas, de forma a:

- Proporcionar uma experiência que não defraude as expectativas de privacidade que um utilizador tem de um operador de telecomunicações;
- Cumprir os requisitos legais, atualmente em aprovação a nível Europeu, que podem constituir uma vantagem concorrencial perante as empresas *Web* americanas;
- Reduzir a "recompensa" no caso de ataques de segurança, desmotivando os potenciais atacantes de o fazer.

Os operadores tipicamente procuram a *killer-application*, mas para que ela surja, os grandes desafios têm de estar resolvidos. A privacidade é provavelmente o *killer-issue* das comunicações máquina-a-máquina. O ecossistema M2M tem um enorme potencial mas vive com paredes de vidro. Cabe ao operador garantir, e passar a mensagem, que apenas intervenientes autorizados conseguem espreitar a informação gerada pelo novo mundo de milhões de dispositivos ligados.

REFERÊNCIAS

[1] Punie, Yves. "A social and technological view of Ambient Intelligence in Everyday Life: What bends the trend?" Key deliverable, The European Media and Technology in Everyday Life Network (EMTEL), 2003.

[2] Orwell, George. "Nineteen eighty-four." 1949.

[3] McDonald, Aleecia, and Cranor, Lorrie. "Beliefs and behaviors: Internet users' understanding of behavioral advertising." 2010.

[4] Lederer, Scott, et al. "Personal privacy through understanding and action: five pitfalls for designers." 2004.

[5] Barbaro, Michael and Zeller, Tom Jr. "A Face Is Exposed for AOL Searcher No. 4417749." New York Times, 2006. http://select.nytimes.com/gst/abstract.html?res=F10612FC345B0C7A8CDDA10894DE404482

[6] Bilton, Nick and Stelter, Brian. "Sony Says PlayStation Hacker Got Personal Data." New York Times, 2011. http://www.nytimes.com/2011/04/27/technology/27playstation.html

[7] Greenwald, Glenn. "NSA collecting phone records of millions of Verizon customers daily." The Guardian, 2013. http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order

[8] Miller, Claire C. "Tech Companies Concede to Surveillance Program." New York Times, 2013. http://www.nytimes.com/2013/06/08/technology/tech-companies-bristling-concede-to-government-surveillance-efforts.html

[9] Warren, Samuel D. and Brandeis, Louis D. "The Right to Privacy." Harvard Law Review, p. 193-220, nº 4, 1890.

[10] Press conference by Viviane Reding on the Data Protection Proposal, 2012. http://ec.europa.eu/avservices/video/player.cfm?ref=82655&sitelang=en

[11] Schonfeld, Erick. "The Online Ad Recession Is Officially Here: First Quarterly Decline In Revenues." TechCrunch, 2009. http://techcrunch.com/2009/05/01/the-online-ad-recession-is-officially-here-first-quarterly-decline-in-revenues/

[12] Singer, Natasha. "Mapping, and Sharing, the Consumer Genome." New York Times, 2012. http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html},

[13] Solove, Daniel J. "The digital person: Technology and privacy in the information age." NYU Press, 2004. ISBN: 9780814740378

[14] McCarthy, Bede. "Cost of cyber attacks triples in a year." Financial Times, 2013. http://www.ft.com/cms/s/0/bb3fcc90-ab4a-11e2-ac71-00144feabdc0.html

[15] Article 29 Data Protection Working Party. "Opinion 04/2013 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems ('DPIA Template') prepared by Expert Group 2 of the Commission's Smart Grid Task Force." 2013. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp205_en.pdf

[16] Azevedo, Ricardo et al. "Gestão de Identidade e Privacidade em Redes Convergentes: Uma Análise de Potencial." Saber & Fazer Telecomunicações, 2008.

[17] Seabra, Eduardo et al. "Desenvolvimento de um Gestor de Identidades para suporte da tecnologia Microsoft CardSpace." Saber & Fazer Telecomunicações, Vol. 1, 2009.

[18] Gonçalves, João et al. "Partilha Social de Serviços e Informação Pessoais – Uma Perspectiva IdM e Privacidade." Saber & Fazer Telecomunicações, 2011.

[19] Sweeney, Latanya. "Simple demographics often identify people uniquely." Health (San Francisco), 1–34, 2000. http://dataprivacylab.org/projects/identifiability/paper1.pdf

[20] Fung, Benjamin et al. "Privacy-Preserving Data Publishing: A Survey on Recent Developments." ACM Computing Surveys (CSUR), 2010.

[21] Adam, Nabil R. and Worthmann, John C. "Security-control methods for statistical databases: a comparative study." ACM Computing Surveys (CSUR), 1989.

CVS DOS AUTORES

João Miguel Gonçalves licenciou-se em Engenharia Informática pela Universidade de Coimbra em 2006 com média final de 15 valores. No mesmo ano foi admitido como bolseiro no Instituto de Telecomunicações em Aveiro, no âmbito do projeto europeu C-MOBILE. Em Junho de 2008 completou, com Distinção, o Mestrado em Wireless Networks pela Queen Mary University of London. Em Setembro de 2008 foi contratado pela PT Inovação, tendo paralelamente iniciado o programa doutoral MAP-i. A tese de doutoramento, em finalização, foca questões de privacidade em cenários futuros de captação universal de informação, cujos trabalhos se relacionam com os mais recentes projetos europeus em que participou, respetivamente C-Cast e SOCIETIES.

Filipe Cabral Pinto é detentor do grau de doutoramento em telecomunicações pela Queen Mary University of London. É atualmente consultor na PT Inovação onde investiga a integração de sistemas M2M em arquiteturas orientadas ao serviço. O seu trabalho tem-se centrado em sistemas de comunicações móveis, desde a camada de transporte à camada de serviço, focando-se em particular na distribuição inteligente de conteúdos multimédia para grupos de utilizadores em redes de próxima geração. Tem estado envolvido em projetos de investigação europeia desde 2002, destacando-se os projetos NetGate, OPIUM, B-BONE, C-MOBILE, C-CAST, VOICES e, mais recentemente, o projeto IoT.est.

10 Uma Gestão Integrada das Equipas de Operação e da Segurança de Edifícios

PAULA CRAVO PAULO FERRO

PALAVRAS CHAVE
Segura, SIGO®, controlo de acessos, operacionalização de recursos, segurança de edifícios

O artigo introduz o conceito da gestão integrada das equipas de operação, a prestar serviço no terreno, e da necessidade do controlo de acessos físicos no contexto da segurança informática e alinhado com as normas ISO/IEC 27001 e ISO/IEC 27002.

É apresentada uma solução integrada baseada em dois produtos desenvolvidos pela PT Inovação, o SIGO® e o Segura, visando a otimização de recursos num ambiente controlado e seguro.

Apresenta-se um *use case* de integração entre os dois sistemas, evidenciando as sinergias óbvias entre a gestão da força de trabalho e o controlo de acessos a locais, permitindo a gestão de equipas e trabalhos associados e tirando partido da integração com sistemas de circuito fechado de televisão, com deteção de situações anómalas e a correspondente geração de alarmes. O *use case* contempla ainda a integração com outros sistemas de controlo de acessos existentes, permitindo o cadastro e gestão centralizada das permissões de acesso dos utentes de um local ou empresa.

1. INTRODUÇÃO

A competitividade crescente no mercado dos serviços exige das organizações um grande controlo dos seus recursos tanto em termos de custos de utilização como de segurança.

A distribuição dos trabalhos pelos seus colaboradores tem como principal objetivo maximizar os resultados obtidos, tirando partido dos conhecimentos, posicionamento, capacidades e responsabilidades, minimizando os custos de operação, tanto nas equipas internas como na interação com entidades externas.

Este facto, associado à necessidade de saber constantemente o que se passa, obriga ao registo de todas as atividades, servindo inclusive para retirar indicadores de performance que retroalimentam os processos gerando a sua otimização.

Todo este processo passa forçosamente pelo tema da limitação e controlo de acessos a um edifício ou local, sendo uma preocupação que existe desde tempos imemoriais, quer para proteção de pessoas e bens, quer para proteção de informação. Modernamente, os meios de controlo existentes envolvem sistemas eletrónicos e informáticos com elevado grau de complexidade.

Adicionalmente a esta necessidade de controlar quem tem acesso, surge a necessidade de auditar quem acede aos locais. Esta situação é ainda mais pertinente numa organização que tenha obrigação de ceder acesso às suas infraestruturas a entidades externas – quer porque no mesmo local estão instaladas estruturas dessas entidades, quer pela subcontratação de serviços de instalação e suporte a terceiros.

Por tudo isto, a necessidade de controlar eficazmente o que cada um faz, a que local acede, a necessidade de ter registos de quando alguém está/esteve a trabalhar numa determinada tarefa ou num determinado local, e a otimização na avaliação de situações anómalas que necessitam de intervenção local são os principais objetivos da integração entre o SIGO e o Segura, que descrevemos em seguida.

1.1 ORGANIZAÇÃO DO ARTIGO

Este artigo começa por apresentar o SIGO e o Segura, com um enquadramento das normas mais recentes relativas à segurança da informação e ao controlo de acessos físicos a locais e edifícios.

Seguidamente, descreve um *use case* onde o problema existente relativamente à operacionalização dos recursos e ao controlo dos acessos físicos é apresentado. É apresentada uma solução que integra estes dois sistemas de forma a responder ao *use case* descrito.

Finalmente, são apresentadas algumas conclusões sobre as vantagens desta integração, bem como possíveis áreas de atuação que podem igualmente ser otimizadas com a integração dos dois sistemas.

2. SIGO

O SIGO® é um produto que dá resposta à complexidade dos ambientes de gestão operacional das empresas atuais. Incorpora componentes para gestão de ordens de trabalho, gestão de equipas (*Workforce Management*) e gestão de problemas da rede e dos fornecedores, em ambiente *WEB*.

>>>

2.1 ENQUADRAMENTO

O SIGO segue e enquadra-se no modelo eTOM [1] do TM Forum, que define um conjunto de áreas de negócio processuais chave, dispostas numa matriz multicamada, necessárias para gerir de forma eficiente e ágil uma organização.

No diagrama da Figura 1 pode observar-se o posicionamento dos vários módulos SIGO face às áreas processuais definidas por este modelo.

As diferentes áreas de atuação do SIGO dão origem a dois produtos, onde cada tema é abordado de forma separada e específica: SIGO TTK e SIGO WFM.

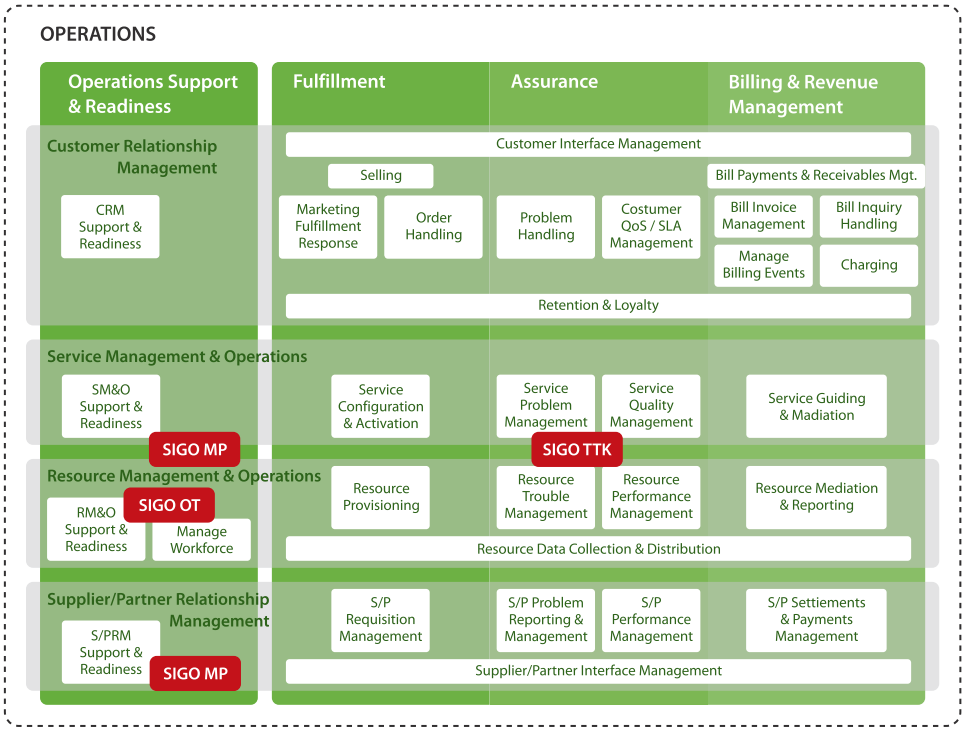


Figura 1. Enquadramento do SIGO no modelo eTOM

2.2 SIGO TTK

O SIGO TTK cobre o tema da gestão de problemas dentro da organização e o respetivo envolvimento com os fornecedores, necessário para resolver os problemas. Esta temática é abordada do ponto de vista de manutenção preventiva e corretiva.

Em termos de manutenção preventiva, o sistema possibilita o carregamento de um plano com rotinas, que especificam os locais/equipamentos a visitar/verificar em determinadas datas, contendo em cada rotina um conjunto de fichas de intervenção onde serão registadas as verificações feitas, sob a forma de *checklists*, e as observações pertinentes para garantir um bom funcionamento de todas as infraestruturas.

Na componente de manutenção corretiva atua-se em reação a algum evento, seja este oriundo de um sistema externo que mantém interface com o SIGO ou de um técnico que deteta um potencial problema que deve ser resolvido. Desde o seu registo até à sua resolução é registada toda a informação dos intervenientes, assim como o que se passava efetivamente, o que foi feito para resolver e como.

Sempre que seja necessário envolver entidades externas com as quais estão celebrados contratos (SLA), existe um módulo dedicado para controlar todo o fluxo de troca de informação, bem como o respetivo cumprimento dos contratos.

Na sequência de todas as intervenções feitas, as unidades em falha que seja necessário substituir são geridas por um módulo próprio que acompanha todo o ciclo de reparação/substituição, permitindo manter registos de todo o percurso de cada unidade, com impacto direto na gestão de *stock*.

Estas funcionalidades encontram-se representadas no diagrama da Figura 2.

>>>

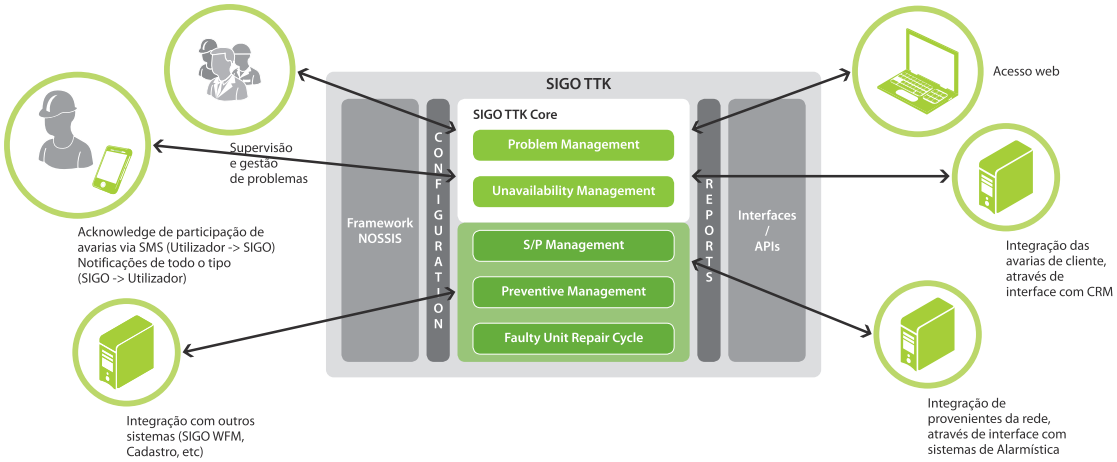


Figura 2. Funcionalidades do SIGO TTK

2.3 SIGO WFM

O SIGO WFM incide sobre a área da gestão das pessoas necessárias para desempenhar todo o tipo de tarefas dentro da organização.

São constituídas equipas de trabalho em função do agrupamento de valências dos seus membros nas diversas áreas. Um elemento pode pertencer a mais que uma equipa de trabalho.

Todo o trabalho realizado na organização é registado em ordens de trabalho que apresentam três fases principais:

1. Planeamento
2. Execução
3. Verificação

Em cada uma das fases são envolvidas as equipas certas para cada tarefa e cada interveniente regista toda a informação relevante para o desenrolar do seu trabalho e dos restantes colegas envolvidos.

O sistema disponibiliza um meio de controlo de todas as tarefas desempenhadas por todos os elementos da organização de forma centralizada, independentemente da origem e vocação das mesmas.

Sempre que existe necessidade de acesso a locais, é registada a entrada e saída acompanhada de mais informação que seja relevante para todos os trabalhos em curso.

Estas funcionalidades encontram-se representadas no diagrama da Figura 3.

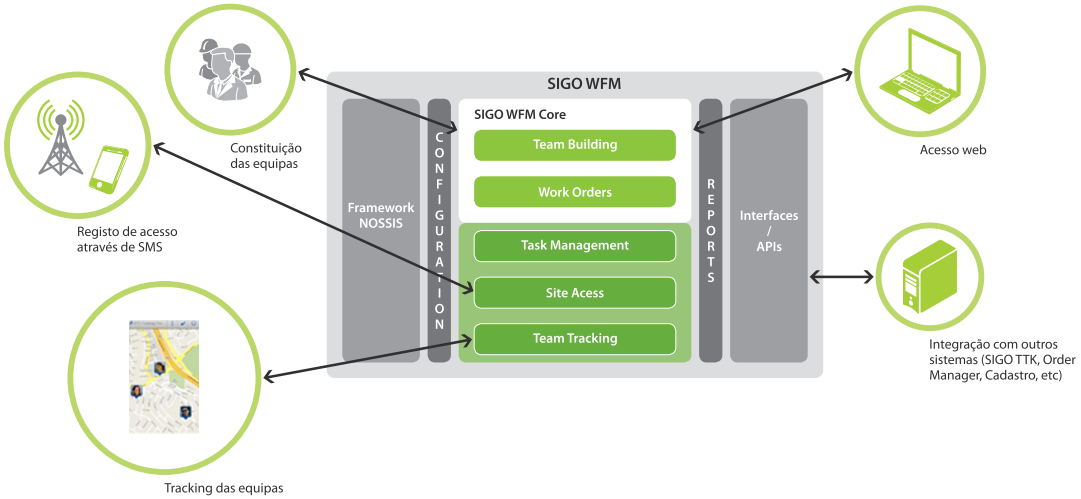


Figura 3. Funcionalidades SIGO WFM

2.4 TRANVERSAL/COMUM

É possível em qualquer altura relacionar os diversos casos, agilizando a sua gestão independente ou em conjunto.

Um completo sistema de notificações mantém todos os interessados a par das alterações mais relevantes, seja por *email*, *sms* ou pelo próprio interface gráfico, recorrendo a alertas intrusivos ou meramente informativos.

A disponibilização de APIs públicas permite a interação com todo o tipo de outras ferramentas da organização, facilitando a troca de informação e construção de processos E2E (*End-to-End*).

Um dos pontos fortes do sistema é a grande capacidade de configuração que é oferecida ao cliente final, permitindo uma total adaptação aos processos existentes, sem deixar de conduzir à adoção de melhores práticas que garantem um maior aproveitamento dos recursos.

3. SEGURA

O Segura é um produto que integra a gestão e controlo de acessos de pessoas a locais. Esta solução inclui ainda a gestão da alarmística associada à segurança de locais e o controlo de equipamentos existentes (portões automatizados, AC, etc.), sendo também possível integrar com soluções de videovigilância.

3.1 ENQUADRAMENTO

As normas da série ISO/IEC 27000 [2] fornecem um conjunto de recomendações e boas práticas referentes à gestão da segurança da informação, riscos e controlo no contexto de um sistema de gestão para a segurança da informação - ISMS. Esta série está a ser revista, sendo que a mais recente atualização inclui a substituição da norma ISO/IEC 27001:2005 pela ISO/IEC 27001:2013.

Esta norma inclui requisitos respeitantes à segurança dos recursos humanos, i.e., controlos aplicados antes, durante e após o período de emprego, controlo e gestão de acessos dos utilizadores, segurança física dos edifícios da organização e do equipamento, segurança operacional e gestão de incidentes, entre outros aspetos.

Como exemplo do âmbito de aplicação desta norma, no que diz respeito à segurança física e ambiental, podemos referir:

- Monitorização do acesso físico aos locais e edifícios, bem como a qualquer infraestrutura de suporte
- Revisão e aprovação periódica das listas de pessoas autorizadas a aceder a determinadas áreas
- Instalação de câmaras de vigilância em áreas críticas e monitorização das imagens por pessoal certificado
- Registo de data e hora de entrada e saída do pessoal e de visitantes (juntamente com o registo do motivo da vista)
- Colocação, de forma visível, do identificador de cada pessoa e apresentação do mesmo para inspeção quando solicitado.

Os exemplos anteriores são meramente ilustrativos. A norma não limita a definição das políticas de acesso nem impõe a sua aplicação.

3.2 CONTROLO DE ACESSOS FÍSICOS

O controlo de acessos físicos [3] diz respeito à restrição de acesso de entrada a uma propriedade, edifício ou sala apenas a pessoas autorizadas.

Um sistema de controlo de acessos físicos determina quem, onde e quando está autorizado a entrar ou sair de um local. Os meios eletrónicos de controlo de acessos (ex. torniquetes, trincos eletrónicos, etc.) são mais versáteis que os meios mecânicos porque não implicam a mudança da fechadura em caso de perda da chave, disponibilizam formas de controlo de acesso mais variadas dependendo da sua complexidade (ex. cartão eletrónico, leitura de palma da mão, leitura de retina, introdução de código, etc.), não estão sujeitos a falhas humanas, e permitem o registo automático de entradas e saídas. Além disso, quando corretamente configurados e integrados num sistema de gestão de segurança, permitem detetar situações anómalas e ativar procedimentos adequados, quer de forma automática (ex. um detetor de fumo automaticamente abre todas as portas que conduzem a uma saída de emergência), quer ativando alarmes mais graves (ex. um detetor de movimento é ativado num local de acesso restrito sem que haja registo prévio de entrada nesse local e um alarme de intrusão é ativado em consequência).
O sistema de controlo de acessos e alarmística Segura (Figura 4), desenvolvido na PT Inovação, pretende dar resposta aos problemas de segurança nos locais e edifícios da PT de forma centralizada, integrando um sistema de alarmística própria e integrando com outros sistemas existentes de alarmística e videovigilância, de forma a fornecer informação em tempo real a uma central de operação.

>>>

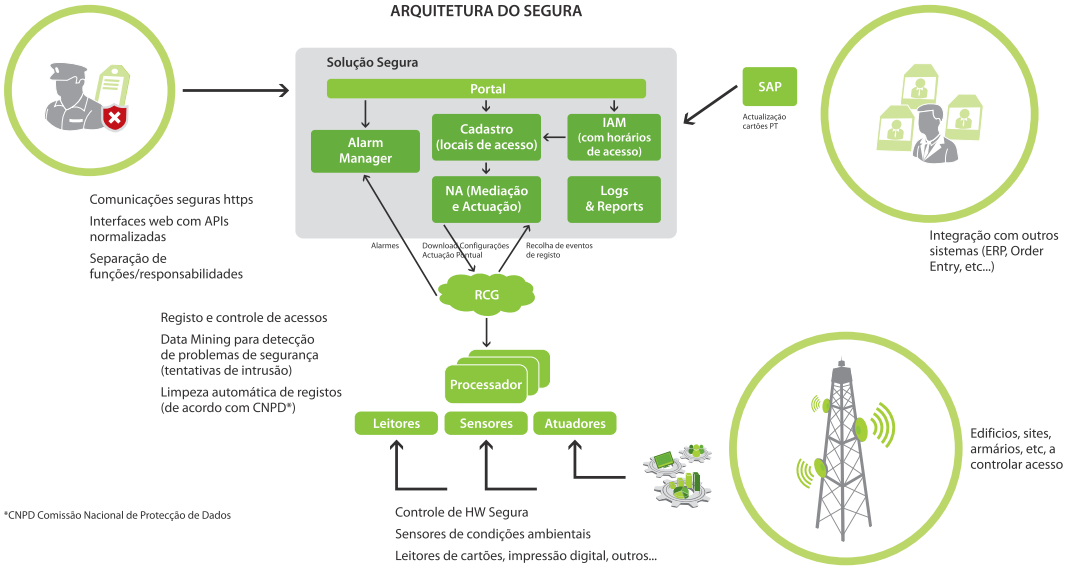


Figura 4. Sistema Segura

O sistema Segura tem por objetivo o controlo de acessos a edifícios PT. Os colaboradores PT e os colaboradores externos têm cartões de identificação, previamente cadastrados na aplicação. Para as visitas existem disponíveis cartões de identificação de visitante que, no registo de entrada, serão associados provisoriamente a uma pessoa. O sistema Segura regista os acessos e mostra alarmes referentes a várias situações irregulares nos acessos.

Em cada edifício é instalado um conjunto de equipamentos de controlo de periféricos (leitores, sensores, relés) para controle das portas de acesso. Este controlo de acesso funciona *off-line*, i.e., sem estar em comunicação com o sistema de gestão. Um operador tem a possibilidade de abrir portas mesmo sem a apresentação de um cartão válido, nos seguintes casos:

- por ordem dada pelo utilizador do sistema, que possua permissões para tal;
- em situações de emergência, a partir de uma chave física ou chave mestra.

O sistema regista todas as entradas e saídas, bem como autorizações de acesso, que ficam acessíveis no sistema durante um período de 180 dias. Além destes registos, o sistema também deteta e regista situações de alarme ou problemas, nomeadamente problemas nos equipamentos (cartas e equipamentos terminais) ou problemas de rede.

O sistema Segura possui *hardware* descentralizado, exemplificado na Figura 5, com autonomia própria, e capaz de manter uma elevada quantidade de informação quando não se encontra em comunicação com o sistema de gestão central. Além disso,

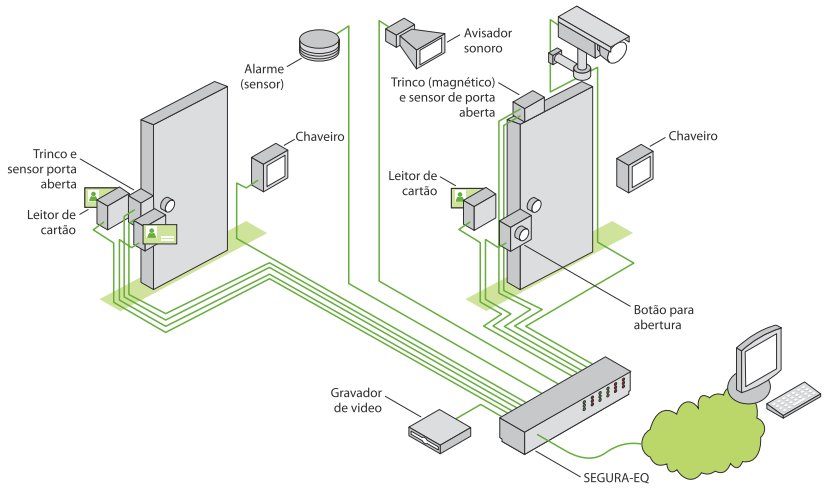


Figura 5. Instalação do Sistema Segura: 2 portas; 1 câmara; um equipamento autónomo SEGURA-EQ; 1 sistema de Gestão

>>>

faz uma gestão integrada de alarmes, processa eventos e permite a correlação de alarmes. Está também desenhado para integrar sistemas de videovigilância e de CCTV (*Closed Circuit Television*), possibilitando o controlo das câmaras e a gravação de imagens em determinadas situações.

A Figura 6 exemplifica o Sistema Segura como solução integradora dos sistemas de controlo de acessos, alarmística e videovigilância da empresa.

A solução Segura permite integrar sistemas de controlo de acessos legados, bem como sistemas de videovigilância existentes, desde que estes disponibilizem uma API de comunicação. Além disso, permite a comunicação com outros sistemas de alarmística, recebendo alarmes destes ou enviando alarmes para outro sistema central.

Toda a informação de pessoas e de utilizadores pode ser obtida de *stores* de utilizadores empresariais e bases de dados de colaboradores (ERP).

Os pedidos de acesso podem ser recebidos no sistema de pedidos (*Order Entry*), que depois de aprovados serão enviados para o Sistema Segura onde o acesso é configurado. Esta configuração de acessos pode ser automática, gerada por um *workflow* de configurações, ou manual, comandada por um operador do sistema.

4. CASO DE ESTUDO

No âmbito da gestão de equipas no terreno, da responsabilidade do SIGO, todas as intervenções que requerem autorização de acesso a instalações controladas pelo Segura têm vantagens adicionais num cenário de integração. Exemplos dessas intervenções são: atividades de reparação (TTKs); atividades de instalação (construção de rede ou provisão de serviço); manutenções preventivas; etc. De seguida apresentamos o cenário de manutenção preventiva para ilustrar as suas vantagens.

4.1 CENÁRIO

Na sequência das manutenções preventivas a um *site* (local) que determinada equipa (conjunto de pessoas) terá que desempenhar ao longo de um período de tempo pré-definido, é necessário garantir que todo o trabalho é registado e que os intervenientes têm as autorizações de acesso necessárias, permitindo o controlo de todo o processo e posterior auditoria.

4.2 SOLUÇÃO

No início de cada ano de trabalho na organização, são planeadas, autorizadas e carregadas em SIGO TTK (módulo MP) as visitas aos locais para efetuar atividades de manutenção preventiva, em que são envolvidas as equipas e seus membros de acordo com as respetivas valências definidas no SIGO WFM.

O carregamento das rotinas, uma vez que pressupõem autorização para a execução dos trabalhos, despoleta os respetivos registos de autorização no Segura, para as pessoas e períodos estipulados.

Chegada a hora de executar o processo de manutenção preventiva, os técnicos deslocam-se ao local onde têm que registar as verificações contidas nas fichas da rotina. Ao passarem o cartão de identificação no leitor de cartões, este valida a autorização para acesso e abre a porta.

Este evento despoleta uma comunicação com o SIGO WFM na componente de acessos a *sites*, onde regista a hora de entrada do técnico identificado e informação dos vários sensores existentes no local, como por exemplo humidade e temperatura.

Esta informação é usada em todo o SIGO e apresentada a todos os utilizadores que estão a criar novas tarefas ou gerir as existentes relacionadas com o local de acesso, para que tenham conhecimento da permanência de um determinado técnico num

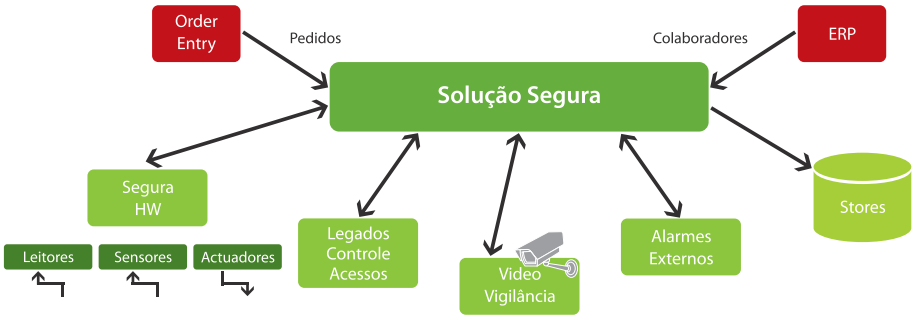


Figura 6. Solução Integradora

determinado local. Tal permite uma melhor gestão das equipas, evitando a deslocação de outra equipa ou técnico ao mesmo local para atendimento de um novo caso, desde que as valências sejam as adequadas.

4.3 OUTROS CENÁRIOS

A título de exemplo, descrevem-se brevemente em seguida alguns cenários que podem ser satisfeitos com esta solução de integração:

- Numa intervenção que carece de autorização superior, pela criticidade que apresenta e/ou impacto das indisponibilidades de serviços em causa, o acesso ao respetivo *site* deve ser controlado, como garantia que a operação só é realizada depois de reunidas todas as condições/autorizações.
  O sistema Segura possibilita o controlo de acesso nestas situações: por comando do operador do sistema Segura, que autoriza e regista o acesso extraordinário ao *site*, atuando sobre os comandos de abertura de porta, a equipa de intervenção tem acesso ao *site*. Estes comandos e o motivo da abertura de porta pelo operador do sistema ficam registados e podem ser auditados.

- Na sequência da conjugação de vários alarmes relacionados com um *site* (temperatura alta, humidade baixa, equipamentos em falha), o técnico da supervisão deve ter a capacidade de fazer uma primeira validação. Se o *site* em causa tiver uma instalação do CCTV integrada no sistema Segura, o técnico pode aceder ao sistema e, controlando as câmaras existentes, observar o local reunindo o máximo de informação que permita tomar a decisão das equipas certas a chamar ao local.

- Cruzar a informação dos acessos com as das intervenções no sentido de detetar possíveis situações de abuso/fraude.

5. CONCLUSÕES

Com esta abordagem, toda a informação enviada pelo Segura pode servir para melhorar a gestão dos trabalhos em curso, solicitando de imediato ao técnico para efetuar outra tarefa aproveitando a mesma deslocação, ou mesmo servir de justificação para uma avaria despoletada por alarmística e que fora provocada inadvertidamente pelo técnico no terreno, facilitando a sua resolução e o retomar do normal funcionamento.

A integração entre estes dois sistemas permite que a informação flua entre ambos sem que os próprios intervenientes se apercebam e tenham que se preocupar com isso, limitando-se a executar o trabalho técnico para o qual foram destacados.

É evidente a redução de custos que todo o processo apresenta, bem como o controlo de acessos onde tudo fica registado, permitindo uma auditoria aos dados e cruzamento com outras fontes de informação.

Toda esta informação é uma fonte de indicadores muito forte que pode levar à otimização de processos.

Esta integração apresenta vantagens face aos cenários hoje existentes, nomeadamente: garante a automatização das autorizações de acesso de equipas no terreno aos locais de intervenção; monitora e regista de forma coordenada e automática os acessos às instalações pelas equipas; complementa a informação das intervenções com dados adicionais (sensores, alarmes, vídeos) permitindo a deteção e correção de situações anómalas.

REFERÊNCIAS

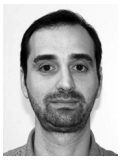
[1] http://www.tmforum.org/BusinessProcessFramework/1647/home.html
[2] http://www.27000.org/index.htm
[3] http://en.wikipedia.org/wiki/Access_control

CVS DOS AUTORES

Paula Cravo, MSC em Information Technology - Information Security pelo Carnegie Mellon University e Mestre em Segurança Informática pela Faculdade de Ciências da Universidade de Lisboa, em 2011, Licenciada em Engenharia Eletrónica e Telecomunicações pela Universidade de Aveiro, em 1993. Iniciou a atividade profissional no CET em 1992, na área de Desenvolvimento de Produtos, tendo estado envolvida no desenvolvimento de aplicações para gestão local de diversos tipos de equipamentos de transmissão: MUXFLEX, SERTO2000, PEDEL, etc. De 1998 a 2001 a sua área de atividade centrou-se no desenvolvimento de aplicações de gestão SNMP para o sistema AGORA. De 2001 a 2005 desenvolveu agentes SNMP para sistemas de transmissão da PT Inovação. No período entre 2006 e 2010 fez parte da equipa de desenvolvimento do sistema ArQoS de testes, monitorização e recolha de indicadores de QoS para diversos tipos de tecnologia de rede (fixa/móvel/IP). Atualmente encontra-se a trabalhar na área de gestão de identidades e controlo de acessos aplicacionais e físicos.

Paulo Ferro, Licenciado em Engenharia Informática pela Faculdade de Ciências e Tecnologia da Universidade de Coimbra, em 1997. Iniciou atividade profissional no CET em 1997, na área de Serviços de Engenharia, tendo estado envolvido no desenvolvimento de Sistemas de Informação Geográfica ligados à representação de redes de telecomunicações, fazendo uso de tecnologias de Base de Dados e WEB. Em 2000 vincou a sua área de atuação nos sistemas de cadastro de redes, fazendo o elo de ligação com a área dos SIGs. Em 2003 integrou a equipa de desenvolvimento de soluções para Gestão de Operações, até à atualidade, período durante o qual se deu origem ao SIGO.

11 Projeto MobiPag


JOSÉ BONNET LUIS CORTESÃO RICARDO MELO

PALAVRAS CHAVE
Pagamentos móveis, NFC, segurança UICC, usabilidade

O projeto MobiPag propôs-se desenvolver um ecossistema de soluções e serviços de valor acrescentado sobre pagamentos móveis.

Do ponto de vista tecnológico, destaca-se a utilização do NFC (Near Field Communication) para a comunicação entre os telemóveis dos clientes e os dispositivos de ponto-de-venda (também eles telemóveis, mas com uma aplicação específica). A segurança do sistema é garantida por uma infraestrutura global que recorre ao SIM como o elemento seguro. O ecossistema de aplicações liga-se a um sistema de back-end com funções de validação e interface com o sistema bancário e com o sistema dum fornecedor de serviços, desenvolvido exclusivamente para o piloto.

O projeto MobiPag demonstrou, no piloto realizado, em ambiente real mas controlado, a viabilidade técnica da abordagem e tecnologias escolhidas, bem como a experiência do utilizador final. Esta validação decorreu no campus da Universidade do Minho, em Guimarães, durante um mês, e envolveu 16 utilizadores 'clientes' e 4 'comerciantes', em 6 cenários de aquisição e utilização de bilhetes (de cantina e autocarro), bem como de receção e utilização de cupões de desconto.

O grande fator distintivo do projeto MobiPag, face a outros projetos e iniciativas na área do dinheiro digital, está nos serviços inovadores que podem ser desenvolvidos tendo por base o pagamento. Mais do que criar um meio de pagamento novo, o objetivo é explorar serviços de utilidade e conveniência para o consumidor e para os comerciantes, que podem ser desenvolvidos em torno da transação de pagamento.

1. INTRODUÇÃO

O projeto MobiPag - Iniciativa Nacional para Pagamentos Móveis - foi pensado com a finalidade de criar uma solução tecnológica para pagamentos móveis em Portugal, aberta e dinamizadora de novos serviços e soluções para o mercado. Nesse sentido, foi desenvolvida uma plataforma aberta que suporta a criação de um ecossistema de soluções e de serviços de valor acrescentado sobre a transação de pagamento.

Em complemento, foram efetuadas sessões do piloto demonstrador no campus de Guimarães da Universidade do Minho, realizado em ambiente real, embora controlado (3 sessões, num total de 16 'clientes' e 4 'comerciantes'), envolvendo alunos, docentes e investigadores, onde se realizaram diversas ações de pagamento - compra de bens e serviços nos bares, compra e validação de senhas de refeição, compra e validação de títulos de transporte e obtenção e utilização de cupões de oferta, recorrendo ao uso de telemóveis com tecnologia NFC e com a aplicação MobiPag instalada.

Neste artigo propomo-nos descrever a arquitetura da plataforma MobiPag, os seus diversos componentes – *applet, middleware, apps* (*user* e *merchant*), *backend* e *service provider* – detalhando as suas principais características, o modo como permitem suportar serviços de valor acrescentado sobre a transação de pagamento e, em particular, o protocolo NFC-*based* associado. Abordaremos igualmente os requisitos de *deployment* da solução, o interface com instituições de pagamento e as questões de segurança associadas (nomeadamente a utilização do SIM – *Subscriber Identity Module* – como *Secure Element*).

O projeto MobiPag foi financiado pelo AdI/Compete e resulta de um consórcio de que fazem parte a CardMobili, líder do projeto, CreativeSystems, Multicert, PT Inovação e Wintouch, bem como a Universidade do Minho, Instituto Superior Técnico e Faculdade de Engenharia da Universidade do Porto. O apoio burocrático da Iniciativa Nacional para Pagamentos Móveis esteve a cargo do Centro de Excelência em Desmaterialização de Transações (CEDT) e contou com um *Advisory Board* constituído por entidades da área financeira (SIBS, Caixa Geral de Depósitos, BES, Millennium BCP, BPI, Visa Europa e MasterCard Europa), operadores de telecomunicações móveis (TMN, Vodafone e Optimus) e empresas da área tecnológica (CTT).

2. PLATAFORMA MOBIPAG

A plataforma de pagamentos móveis MobiPag consiste num conjunto de componentes de software que implementam o processamento de transações de pagamento com equipamentos móveis, utilizando tecnologia NFC. Adicionalmente, e sem perda de generalização, a presente arquitetura foi desenhada na assunção de que o elemento seguro, presente em ambos os dispositivos envolvidos numa transação, está dentro do UICC, e que os dispositivos usam Android como sistema operativo. A arquitetura assume que nenhum dos dispositivos, de cliente ou de comerciante, são entidades confiáveis (dado que os equipamentos e os canais de comunicação podem ser corrompidos e/ou interceptados), mas considera que os serviços centrais MobiPag e que o código que corre nos elementos seguros dentro dos dispositivos são confiáveis.

O projeto desenvolveu tecnologia que permite:

- O registo e armazenamento de diversos métodos de pagamento (vCards) no dispositivo móvel, que serão acedidos por aplicações autorizadas;
- O acesso das aplicações aos métodos de pagamento, o que possibilita a criação de soluções em diversas áreas de negócio, como a bilhética, a fidelização de clientes, os transportes, etc., adicionando valor ao processo de pagamento;
- O armazenamento seguro dos elementos essenciais aos serviços de valor acrescentado (vTokens), expandindo o nível de segurança até elementos além da transação de pagamento;
- A fácil incorporação de soluções de terceiros, que utilizam os componentes tecnológicos disponibilizados para oferecer outros serviços de pagamento.

ARQUITETURA E COMPONENTES

A arquitetura MobiPag implementa um serviço de pagamentos usando uma multiplicidade de contas financeiras, locais (no caso das eletrónicas) e/ou remotas (no caso das bancárias), servindo para estas últimas como interface entre as lógicas de pagamentos e as instituições financeiras. Adicionalmente, a arquitetura MobiPag foi desenhada de forma a fornecer não só serviços de pagamento, mas também outras lógicas de valor acrescentado, tais como tickets de fidelização ou bilhética, através dos fornecedores de serviço associados.

O serviço de pagamentos fornecido pela arquitetura assume que todos os pagamentos são efetuados de forma online com as entidades financeiras que detém as contas financeiras envolvidas, ou seja, não estão previstas transferências offline de valores entre clientes e fornecedores. A arquitetura assume ainda que em todas as transações são devidamente validados os saldos das contas envolvidas.

A arquitetura assume que clientes e comerciantes comunicam entre si através de NFC, e que são os comerciantes que detém a responsabilidade da efetiva comunicação com as instituições financeiras, feita através de IP e intermediada pela plataforma MobiPag.

Sem perda de generalidade, a arquitetura desenhada prevê que os elementos seguros em cada dispositivo estejam contidos no UICC, e que os dispositivos usam Android como sistema operativo (ver a Figura 1).

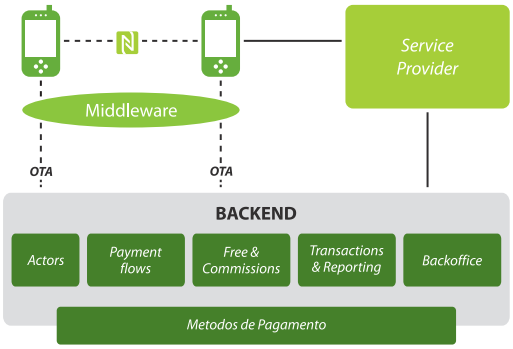


Figura 1. Arquitetura MobiPag de alto nível

Relativamente aos dispositivos móveis, a Figura 2 ilustra os focos de trabalho efetuados no âmbito do projeto, em particular, as Apps de pagamento, a App MobiPag (serviço Android com o Middleware MobiPag) e a Applet instalada no elemento seguro.

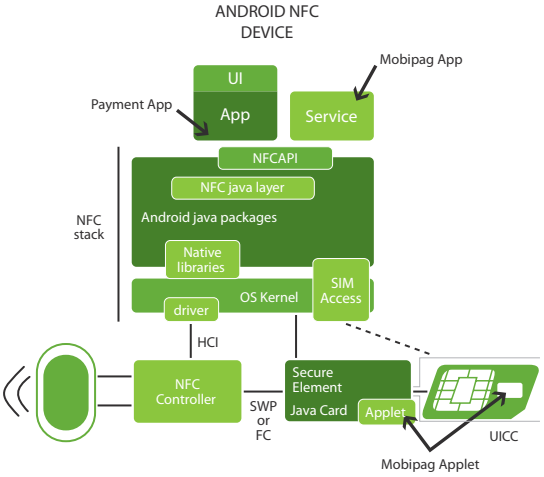


Figura 2. Focos de trabalho efetuados no âmbito do projeto MobiPag

PROTOCOLO

A comunicação entre o *Middleware* MobiPag que corre em cada um dos dispositivos envolvidos numa transação é feita através de NFC, enquanto que a comunicação entre o Middleware do dispositivo do comerciante e a plataforma *Backend* é feita através de WiFi ou HSDPA usando uma simples chamada por https.

O protocolo definido consiste em 4 mensagens trocadas por NFC entre os dispositivos móveis e 2 mensagens trocadas entre o dispositivo do comerciante e a plataforma *Backend*, sendo que estas últimas podem não existir em certas situações como, por exemplo, no caso dos pagamentos offline. A Figura 3 ilustra as mensagens do protocolo.

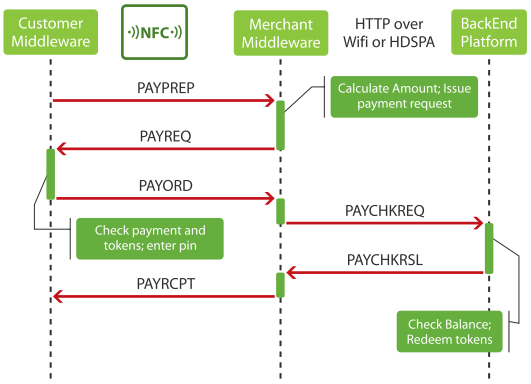


Figura 3. Mensagens do protocolo trocadas por NFC

As mensagens contêm identificadores de transação (TID) para cada uma das 3 entidades participantes no protocolo: cliente (CTID), comerciante (MTID) e plataforma (PTID). Cada um desses identificadores contém: a identificação do participante, um número de transação, o *timestamp* da transação e um *nonce* (em segurança, é um conjunto de dígitos utilizado uma única vez) aleatório.

CTID = {CustomerID, CustomerTransNumber, CustomerTime, CNonce}
MTID = {MerchantID, MerchantTransNumber, MerchantTime, MNonce}
PTID = {PlatformID, PlatformTransNumber, PlatformTime, PNonce}

A primeira mensagem trocada entre o cliente e o comerciante é a mensagem de preparação de pagamento:

PAYPREP = {CTID, vCard, vTokensList, SIG_customer}, Customer_CERT

Além do CTID, a mensagem contém o vCard do cliente, os *tokens* a utilizar (e.g. *tokens* de fidelização, bilhetes), e uma assinatura sobre os 3 items anteriores. Adicionalmente, segue na mensagem o certificado do cliente, de forma a que o comerciante consiga verificar a assinatura.

Após a recepção da mensagem PAYPREP, a aplicação do comerciante irá calcular qual o valor a pagar e/ou quais os *tokens* a abater, e fará esse mesmo abate ou enviará a informação ao Service Provider para que este proceda às operações necessárias. De seguida, constrói um elemento denominado PAYDESC que contém: a denominação do que vai ser adquirido, o identificador da aplicação de pagamento, o custo original, o valor do desconto, e a lista de *tokens* a abater.

PAYDESC = {Goods Description, Payment App, Original Amount, Discount Amount, vTokens-List}

É assim possível construir a mensagem que irá retornar ao dispositivo do cliente, denominada PAYREQ, que contém os TID de cliente e comerciante, o vCard do cliente, um código de status da operação, e uma assinatura sobre estes items. Tal como na mensagem anterior, seguirá o certificado do comerciante, para que o cliente consiga verificar a assinatura.

PAYREQ = {MTID, CTID, PAYDESC, vCard, status, SIG_merchant}, Merchant_CERT

Os *timestamps* dentro dos TID de cliente e comerciante tem uma diferença máxima admissível, que caso seja ultrapassada leva ao cancelamento da transação.

A mensagem seguinte é denominada PAYORD, enviada pelo cliente para o comerciante para que seja confirmado o pagamento. Esta mensagem contém essencialmente o mesmo que a mensagem PAYREQ, levando apenas a assinatura do cliente.

PAYORD = {MTID, CTID, PAYDESC, vCard, SIG_customer}

Após receber esta mensagem, o comerciante envia uma mensagem PAYCHKREQ ao *Backend* da plataforma MobiPag, para que esta proceda às operações necessárias junto do Service Provider e/ou das instituições financeiras envolvidas. A mensagem PAYCHKREQ consiste na mensagem PAYORD assinada pelo comerciante. Como a mensagem PAYORD já se encontrava assinada pelo cliente, a mensagem PAYCHKREQ acaba por representar um contrato assinado por ambas as partes.

PAYCHKREQ = {PAYORD, vCardMerchant, SIG_Merchant}, [Customer_CERT, Merchant_CERT]

Opcionalmente, os certificados de cliente e comerciante são enviados com a mensagem PAYCHKREQ, apesar de o *Backend* já os poder ter registados.
Ao receber a mensagem PAYCHKREQ, o *Backend* procede às operações necessárias e retorna depois uma mensagem PAYCHKRSL. Nesta mensagem é devolvida a mensagem PAYCHKREQ original, juntamente com um código de status, o TID da plataforma, uma lista de *tokens* (comprados e/ou ganhos durante a operação) e a assinatura sobre estes items. Adicional e opcionalmente, o certificado da plataforma é enviado com a mensagem.

PAYCHKRSL = {PAYCHKREQ, Status, PTID, vTokens-AuthList, SIG_srv}, [Platform_CERT]

Por fim, a mensagem PAYRCPT é enviada pelo comerciante ao cliente, sendo que esta mensagem é uma cópia exacta da mensagem PAYCHKRSL.

PAYRCPT = PAYCHKRSL

De uma forma diferente da mostrada na Figura 3, a figura seguinte ilustra a informação trocada entre o cliente e o comerciante, em cada um dos "taps" NFC (aproximações ou toques entre os dois dispositivos que pretendem comunicar) efetuados.

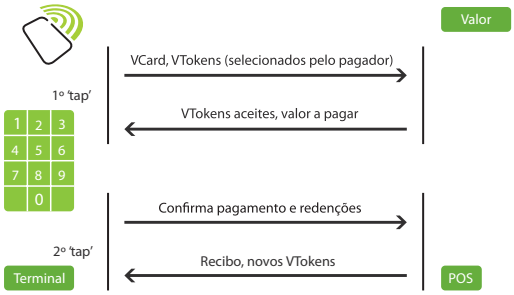


Figura 3. Informação trocada entre o cliente e o comerciante

PRINCIPAIS DIFICULDADES E LIMITAÇÕES

Uma das principais dificuldades encontradas no desenvolvimento deste projeto foi o facto do sistema NFC não permitir a comunicação bidirecional num único "tap". Esta limitação traduzia-se numa experiência de utilizador (*user experience*) pouco apelativa, dado que para a transmissão das 4 mensagens seriam necessários outros tantos "taps". Para dirimir este problema, foi necessário efetuar uma alteração diretamente no *kernel* do sistema operativo dos dispositivos, após a qual passou a ser possível a desejada comunicação bidirecional.

A API de acesso ao SIM card revelou-se outra limitação encontrada, pelo simples fato de a mesma ser muito recente e estar implementada apenas para um conjunto muito limitado de equipamentos.

Por fim, há a salientar que a atual geração de SIM cards ainda revela um baixo desempenho no acesso aos mesmos, o que impactou severamente na velocidade com que as transações eram executadas. A solução passou por compensar esse baixo desempenho com uma redução no tamanho das chaves de segurança utilizadas (e consequente redução do tempo de processamento das mesmas pelos algoritmos de segurança).

3. PILOTO MOBIPAG

Para o piloto do projeto foram escolhidos cenários de pagamento em que produtos e serviços móveis que implementam soluções de pagamento por equipamento móvel, em diferentes sectores de atividade, utilizando a plataforma de pagamentos móveis, fossem testados em ambiente real controlado.

CENÁRIOS DO PILOTO

Através do piloto foi possível, em primeira mão, contactar com as mais recentes inovações em pagamentos com telemóveis, recorrendo a quatro momentos diferentes: pagamento de bens e serviços nos bares, compra e validação de senhas de refeição, compra e validação de títulos de transporte e obtenção e utilização de cupões de oferta. Com caraterísticas distintas no que diz respeito ao pagamento, as demonstrações decorreram em diferentes pontos de venda da Universidade do Minho, de modo avaliar a flexibilidade da plataforma em diferentes ambientes.

Para o efeito, um grupo de 18 participantes voluntários, alunos da academia, realizou diversas ações de pagamento, recorrendo ao uso de telemóveis com tecnologia NFC e com a aplicação MobiPag instalada. Pretendeu-se assim conhecer as funcionalidades e vantagens da solução, validar todas as componentes técnicas, concretamente ao nível da usabilidade, e perceber quais as principais dificuldades apresentadas pelos utilizadores. Foram analisados vários indicadores, como o tempo que demoram a realizar uma operação, se a seleção que fizeram na aplicação corresponde à ação que pretendiam e se os menus apresentados são percetíveis, entre outras questões.

APLICAÇÕES

No âmbito do projeto foram desenhadas e implementadas duas aplicações referência, uma de 'cliente' e outra de 'comerciante', que serviram para a execução dos cenários desenhados, cujos ecrãs iniciais se mostram nas Figura 5 e Figura 6.

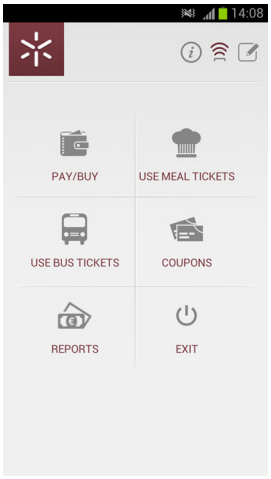


Figura 5. O ecrã inicial da aplicação do Cliente

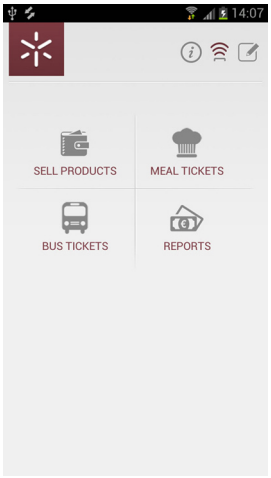


Figura 6. O ecrã inicial da aplicação do Comerciante

RESULTADOS DO PILOTO E CONCLUSÕES

Os resultados do piloto MobiPag evidenciam que as tecnologias e serviços funcionaram de forma integrada e muito aceitável.

No que diz respeito à avaliação da experiência de utilização, a generalidade dos participantes manifestaram uma atitude muito positiva relativamente à utilização da tecnologia usada. Os resultados do piloto demonstram que há um conjunto de elementos que têm uma forte influência no resultado final da experiência de utilização, e que têm sobretudo a ver com a forma. Ao utilizar-se uma tecnologia de comunicação à distância (curta) como é o NFC, é importante que os utilizadores executem a interação correspondente à transação de forma confortável (por exemplo, não necessitando de estar muito tempo com os dois dispositivos próximos, ou ouvindo claramente o som de retorno quando a transação se efetua com sucesso, mesmo em ambientes ruidosos) e segura (por exemplo, na introdução do PIN de validação). Estes fatores têm um grande impacto na forma como todo o sistema de pagamentos é percebido.

A usabilidade associada ao interface implementado é uma componente relevante na perceção de qualidade por quem utiliza este tipo de tecnologias. A escolha adequada da terminologia, imagem de botões e símbolos usados, e o retorno nas ações executadas (sonoro ou por vibração, por exemplo) influenciam determinantemente o desempenho geral na utilização do sistema. Os cenários de validação de bilhetes foram penalizados por limitações da tecnologia utilizada, a que acresceram decisões tomadas ao nível do interface das aplicações. Especificamente, a questão centrou-se no desenho do interface para a aquisição dos bilhetes, que obrigou a um excessivo número de passos distintos necessários.

O facto de o sistema pedir um PIN de confirmação para algumas operações de maior valor mostrou também ser reconfortante para o utilizador, no sentido em que lhe dá confiança na segurança que suporta o sistema.

Concluímos ainda neste estudo que a percepção de valor dos sistemas de pagamentos móveis se centrou sobretudo na possibilidade de integração de bilhetes e cupões de desconto diretamente no processo de pagamento. Foi ainda particularmente apreciada a possibilidade de se comprarem bilhetes não físicos, que são transportados 'no telemóvel', e que depois podem muito facilmente ser validados.

Em termos tecnológicos, a plataforma implementada permite, a todos os operadores interessados, a oportunidade de desenvolver e comercializar produtos e serviços personalizados. A solução cria a oportunidade (e ambição) de suportar modelos de negócios sólidos, permitindo a rápida introdução no mercado de soluções de pagamento passíveis de serem adotadas universalmente, por comerciantes e compradores. Mais do que disponibilizar um novo método de pagamento, o foco incidiu em explorar e habilitar serviços úteis, desenvolvidos em torno de processos de pagamento, aumentando a atratividade da aplicação, a experiência do utilizador e as oportunidades de negócios para os comerciantes. Adicionalmente, as tecnologias utilizadas permitem assegurar a interoperabilidade, segurança, autenticação, ubiquidade e universalidade da solução.

A solução desenvolvida inclui algumas características técnicas que importa relevar:

- Desenho e implementação de um protocolo seguro (2-tap NFC- *based*) capaz de suportar transações de pagamento móvel, mas igualmente esquemas de bilhética e fidelidade;
- Utilização dum domínio de segurança num UICC como elemento seguro numa solução de pagamentos móveis, o que resultou no desenvolvimento dum *applet* que possibilita um acesso controlado às funcionalidades core;
- Combinação do *applet* (a ser executado dentro do elemento seguro) e do *middleware*, como alicerces para todo o ecossistema de aplicações MobiPag;
- Desenvolvimento dum *Service Provider* de referência, demonstrando um cenário *cloud-based* em que "vários" comerciantes partilham o *software* e infraestrutura, mas implementam cada um o seu negócio, de forma completamente independente;
- Desenvolvimento duma plataforma de *backend* como base dum sistema de produção;
- Ligação em tempo real aos sistemas de Qualidade da Caixa Geral de Depósitos, possibilitando a realização duma transação de pagamento de ponto-a-ponto no piloto MobiPag;
- Interoperabilidade transversal: operadoras, meios de pagamento (vcards) e *Service Providers*;
- Modelos colaborativos e de parceria;
- Segurança ponta-a-ponta, baseada no SE do SIM, que abrange todos os elementos da transação;
- Conceito (vtoken) versátil para suportar diversos serviços de valor acrescentado (fidelização, bilhética, cupões, etc.);
- Modelos transacionais baseados em P2P.

Algumas das caraterísticas acima são de particular relevância, dadas as limitações do atual estado da arte da tecnologia, em particular a imaturidade do Open Mobile API e da atual nova geração de cartões SIM, assim como o atraso na divulgação e estabilidade da tecnologia NFC. Apesar do MobiPag ter sido implementado com recurso a estas tecnologias emergentes, não sendo por isso possível a sua imediata massificação, ele pode no entanto ser operacionalizado em ambientes controlados como o do piloto realizado.

A solução MobiPag é caracterizada por uma plataforma de pagamento desacoplado das aplicações de pagamento, com uma solução de gestão de *tokens* capaz de suportar diversos cenários de utilização. Adicionalmente a plataforma permite a associação de *Service Providers* como agregadores de modelos de negócio, adicionando lógica processual específica ao seu negócio, de forma transparente aos desenvolvedores de aplicações e utilizadores finais: os comerciantes e os consumidores.

Na sequência dos trabalhos efectuados, identificou-se um conjunto de desenvolvimentos que permitirão melhorar a solução MobiPag, nomeadamente:

- O modelo de segurança MobiPag foi desenhado para suportar pagamentos NFC-*based*, mas podem também ser utilizados modelos com menos garantias de segurança, permitindo a sua extensão para cenários e-commerce (POS virtuais), com dinheiro real ou virtual sob a forma de *tokens* ou bilhetes;
- Suportar modelos de negócio onde grupos de comerciantes não concorrentes partilham informação de utilização que permita construir perfis de clientes mais ricos (questões de privacidade deverão ser endereçadas);
- Evoluir a plataforma de *backend* para suportar escalabilidade, assim como interoperabilidade entre plataformas de *backend* geridas por outras entidades (eventualmente sediadas em diferentes países);
- Desenhar modelos de negócio sustentáveis, capazes de atrair todos os players do ecossistema.

CVS DOS AUTORES

José Bonnet, licenciado e mestrado (pré-Bolonha) em Engenharia Eletrotécnica e de Computadores (Sistemas Digitais) na Faculdade de Engenharia da Universidade do Porto, liderou a equipa de Sistemas de Informação do Pólo do Porto da PT Inovação entre 1999 e 2008, a equipa de Desenvolvimento de Serviços e Inteligência no Negócio, para os clientes de África Subsariana e Ásia Pacífico, entre 2008 e 2010 e as equipas de Processos de Clientes e Testes Automáticos até Setembro de 2011, liderando depois a equipa responsável pela área de Mobile Money até ao início de 2013. Desde então passou a fazer parte da equipa de Mobile TV.

Luis Cortesão, licenciatura em Engenharia Informática pela Universidade de Coimbra. Como colaborador do CET, PT Inovação e PT Sistemas de Informação, desenvolveu trabalho nas áreas de sistemas de informação geográfica, usabilidade, gestão de competências, gestão do conhecimento, Revenue Assurance e Business Intelligence. Responsável pelo desenvolvimento de sistemas de gestão de fraude. Gestão de produto, desenho e arquitetura de soluções na área de Mobile Money.

Ricardo Melo, licenciado em Engenharia Informática e Computação pela Faculdade de Engenharia da Universidade do Porto, é Engenheiro de Software desde 2003 na PT Inovação, onde começou pela área dos Projetos Europeus IST e Eurescom. Após uma breve passagem pela equipa dos Produtos de Voz, integrou a área de Billing, Redes Inteligentes e Sistemas de Gestão de Clientes, onde colaborou em diversos projetos e para um vasto leque de clientes até 2011. Está desde então ligado à área de Mobile Money, e é atualmente Gestor de Produto na área dos Sistemas de Suporte ao Negócio.

12 VoIP *Peering*




ANTÓNIO AMARAL DAVID GONÇALVES JOSÉ SILVA FRANCISCO FONTES

PALAVRAS CHAVE
IMS, SBC, Interligação

A evolução das redes de dados ao longo dos últimos anos tem sido significativa, verificando-se uma diversificação dos serviços e da forma como são disponibilizados.

Acompanhando o desenvolvimento das tecnologias de telecomunicações, os operadores de voz começaram a disponibilizar os seus serviços sobre redes de dados, sem necessidade de recorrer à rede *Public Switched Telephone Network* (PSTN), através do desenvolvimento de novos métodos e mecanismos. Contudo, as ligações entre operadores ocorrem sobre redes PSTN não sendo possível dar serviços avançados. Para contornar este problema, foram desenvolvidos mecanismos que permitem a comunicação de voz entre diferentes operadores sobre o protocolo *Internet Protocol* (IP), denominados de mecanismos de VoIP *peering*.

O VoIP *peering*, apresenta-se como um mecanismo de encaminhamento de sessões de voz permitindo a conectividade entre diferentes operadores sobre IP. As soluções de VoIP *peering*, caracterizam-se por uma redução de custos, justificada pela não necessidade de encaminhar as chamadas para a rede PSTN, uma melhoriana qualidade da comunicação, porque não é necessário utilizar mecanismos de *transcoding* na rede, e uma otimização dos recursos devido à escolha de rotas com requisitos de QoS (Quality-of-Service) adequados para cada sessão.

A garantia do cumprimento de acordos bilaterais/multilaterais, característica do VoIP *peering*, definidos entre operadores nacionais e internacionais, possibilita a otimização dos lucros através da escolha das rotas com o menor custo possível.

Este artigo apresenta o estado da arte sobre as soluções VoIP *peering*, bem como uma possível implementação da solução numa rede de operador.

1. INTRODUÇÃO

Com o crescimento de soluções VoIP (*Voice over IP*) e com o amadurecimento da arquitetura IMS (IP *Multimedia Subsystem*) [1], os operadores de rede fixa e móvel têm adotado soluções de VoIP *peering*, garantindo assim a disponibilização de serviços multimédia sobre as redes IP entre diferentes operadores, aplicável tanto ao tráfego nacional, como ao tráfego internacional.

Historicamente, o termo VoIP *Peering* é apresentado como um conjunto de práticas levadas a cabo de forma a permitir a comunicação entre fornecedores de serviço sobre o protocolo *Session Initiation Protocol* (SIP) [2].

Através destas, torna-se possível a interoperabilidade a partir do nível de transporte ou de camadas superiores, podendo realizar sessões ponto-a-ponto para os serviços pretendidos e permitindo que os recursos alocados não sejam perdidos.

Inicialmente, o objetivo dos operadores passava por substituir os sistemas de interligação por soluções que reduzissem o OPEX (*Operating Expense*), continuando a garantir os serviços já existentes. Assim, a solução VoIP apresentava-se como uma "Ilha", em que a comunicação interna seria efetuada sobre IP e, para a comunicação com outros operadores nacionais e internacionais, eram utilizadas ligações de comutação de circuitos (PSTN [3]) .

Atualmente, os operadores de telecomunicações disponibilizam soluções ricas ao nível de serviços, necessitando que a interligação com outros operadores seja feita através de soluções de VoIP *peering*, potenciando o negócio à volta destes serviços.

2. ESTADO DA ARTE

Analisando o mercado atual dos operadores de telecomunicações, constata-se que há operadores que já têm as suas redes de voz a operar sobre a rede IP e outros operadores que se encontram no processo de mudança [4]. Em ambos os casos, a interligação entre operadores continua a ser feita maioritariamente sobre a rede PSTN, sendo para isso necessário realizar conversão de pacotes para circuitos.

A interligação de operadores, suportada pelas redes PSTN, permite serviços de voz tradicionais, mas o mesmo não acontece para serviços avançados disponibilizados por soluções convergentes (serviços baseados em sessões SIP), porque não é possível manter a sessão através da rede PSTN.

Enquanto não foi resolvido o problema da interligação entre operadores através da rede IP, a rentabilidade de serviços avançados ficava bastante condicionada, já que não podiam ser disponibilizados esses serviços a clientes fora da rede do próprio operador. Nesse sentido, foram desenvolvidos equipamentos com funções de interligação IP, um dos quais o *Session Border Controllers* (SBCs) [5], que é responsável pelo controlo e gestão de sessões SIP. Desta forma, este equipamento já permite manter sessões SIP entre diferentes operadores, possibilitando a interligação entre redes IP [6].

Tal como referido anteriormente, a interligação entre os operadores pode ser feita ao nível nacional e ao nível internacional. No caso de soluções de VoIP peering, também são consideradas soluções de VoIP *peering* nacional e VoIP *peering* internacional.

PEERING NACIONAL

Ao nível do *peering* nacional, são realizados acordos de interligação entre operadores a prestar serviço no território nacional, sendo estes acordos regulados pela entidade reguladora das telecomunicações. A existência de uma entidade com funções de regulamentação nos cenários de *peering* nacional justifica-se pela necessidade de criar um conjunto de políticas comuns a serem implementadas pelos diferentes operadores, de forma a ser possível a disponibilização de serviços avançados. No caso de Portugal, a entidade reguladora é a ANACOM [8].

No VoIP *peering* nacional, para optimizar o encaminhamento para o operador de destino de números portados, devem ser implementadas soluções de portabilidade numérica. Desta forma, os operadores de origem devem efetuar consultas à base de dados de portabilidade, para verificarem a que operador deve entregar a sessão.

PEERING INTERNACIONAL

O *peering* internacional permite a interligação entre operadores de diferentes países sobre a rede IP. O processo de interligação nacional apresenta-se menos complexo do que o processo de interligação internacional. Ao nível nacional, os operadores encontram-se num espaço geográfico confinado enquanto que ao nível internacional, como a interligação é feita a nível global, existem questões técnicas e económicas que podem complicar mais o processo de interligação.

Para o *peering* internacional é impraticável e desnecessário o estabelecimento de acordos entre todos os operadores, devido ao elevado número e aos diferentes interesses de cada um. A solução passa pelo estabelecimento de acordos de *peering* diretos ou indiretos [7] [2] entre os diferentes operadores.

Os acordos de *peering* diretos ocorrem quando duas entidades se interligam diretamente ao nível da sessão SIP, tipicamente sem custos associados, já que os acordos elaborados partem do princípio que existe um benefício para ambas as partes a nível do serviço obtido [7]. Contudo, pode haver cenários em que tal não se verifica, existindo custos para o operador de origem e/ou para o operador de destino. De referir que neste tipo de acordo, são considerados os três modelos de taxação: *Initiating Party Network Pays* (IPNP), *Receiving Party Network Pays* (RPNP), ou *Bill-and-Keep* (BAK) [3].

Em termos funcionais, as ligações de *peering* direto consideram acordos que definem requisitos técnicos e requisitos de negócio, sendo a interligação da responsabilidade dos elementos envolvidos.

Os acordos de *peering* indireto ocorrem na situação em que existe um operador de trânsito envolvido na interligação de dois operadores que pretendem manter uma sessão SIP. Tipicamente, o operador de trânsito tem relações de confiança com os operadores intervenientes na sessão SIP, sendo o responsável por garantir as condições para que a sessão SIP possa ser estabelecida.

Concretizando num exemplo, considera-se o caso de um utilizador da Timor Telecom, que pretende estabelecer uma sessão com um utilizador da Oi. Se a Timor Telecom não possuir acordos VoIP *peering* com a Oi, mas possuir acordos com a Portugal Telecom e a Portugal Telecom por sua vez possuir acordos com a Oi, então a Portugal Telecom apresenta-se como um possível operador de trânsito entre os dois operadores.

Figura 1. Interligação entre Timor Telecom e Oi, através de *peering* indireto pela PT

Deste modo, os utilizadores da Timor Telecom conseguem estabelecer ligações VoIP com utilizadores da Oi, através dos acordos elaborados entre a Oi e a Portugal Telecom. Para o operador de trânsito, neste exemplo a Portugal Telecom, ficam as receitas associadas com os acordos bilaterais estabelecidos.

Num cenário em que um operador de trânsito tenha diferentes acordos com outros operadores de trânsito, a escolha de qual o operador a usar depende de vários parâmetros associados à existência de requisitos de qualidade de serviço, custos associados à utilização da ligação e a acordos contratualizados entre os operadores aquando do estabelecimento do acordo. O conjunto destes parâmetros tem por base a escolha da rota mais vantajosa a nível de custos para o operador, sendo este tipo de soluções denominadas de soluções de encaminhamento *least-cost routing* (LCR).

MODELOS DE INTERLIGAÇÃO VOIP

A nível de modelos de interligação são considerados dois modelos principais [3], o modelo Bilateral e o modelo Multilateral.

Nos cenários em que os operadores possuem um perfil de tráfego semelhante, tipicamente são estabelecidos acordos seguindo o modelo Bilateral, caracterizado por permitir um ajuste nas políticas estabelecidas, consoante o tipo de acordo que se pretende estabelecer.

Em termos de distribuição das responsabilidades de gestão, o modelo Bilateral apresenta-se como sendo um modelo distribuído, onde os operadores definem as regras de interligação e cada um é responsável pela gestão, configuração e manutenção dos equipamentos de interligação da sua rede.

A vantagem deste modelo está relacionada com o facto de permitir aos operadores intervenientes ter o controlo do acordo, permitindo-lhes definir soluções ao nível de requisitos técnicos e de negócio específicos, garantindo assim uma maior flexibilidade de renegociação dos acordos quando for necessário.

As desvantagens deste modelo, por um lado, estão relacionadas com os custos de implementação, já que é necessário implementar ligações dedicadas entre os dois operadores para garantir requisitos de QoS não podendo, por este motivo, ser interligados sobre a Internet. Por outro lado, o custo da gestão da interligação apresenta-se também como um ponto menos favorável, porque os operadores são responsáveis pela monitoria e resolução de problemas associados à interligação. Este último ponto é particularmente sensível, quando existe a necessidade de alterar com frequência os equipamentos de suporte à implementação do acordo em questão [3].

No modelo multilateral passa a existir uma entidade central responsável pela interligação dos operadores, ou seja, cada operador estabelece um acordo com essa entidade central e a partir daí toda a gestão da interligação passa a ser coordenada por essa entidade.

Por um lado, o facto de existir um "círculo de confiança" [9] para controlo central de todas as funcionalidades de gestão associadas às interligações VoIP, traduz-se numa vantagem deste modelo, onde se garantem os requisitos dos diferentes operadores e facilita a interoperabilidade entre eles através da normalização de protocolos usados para a entrega de chamadas. Por outro lado, a falta de controlo dos operadores nas interligações traduz-se numa desvantagem, porque deixam de ter qualquer controlo no policiamento das ligações, na monitorização de tráfego e no controlo de qualidade de serviço [9].

IP EXCHANGE

IP eXchange (IPX) é um serviço definido pelo *Global System for Mobile Communications Association* (GSMA) [10] e caracteriza-se por permitir a interligação entre diferentes operadores, tendo por base um modelo técnico comum entre os diferentes elementos e possuindo modelos comerciais bem definidos [11]. Providencia um conjunto de especificações técnicas e de negócio, capazes de dar sustentabilidade à disponibilização de serviços IP, e possuindo funcionalidades ao nível da segurança, taxação, interligação e qualidade de serviço.

Ao nível da segurança, o IPX opera sobre uma rede IP privada usando o domínio de *routing* do próprio IPX, encontrando-se separado da Internet.

Ao nível da taxação, existe um controlo da informação de forma a validar o contratualizado e permitindo um conjunto alargado de modelos a serem usados para este efeito. O IPX permite um conjunto diverso de modelos negociados entre as partes intervenientes e podendo ser de diferentes modos conforme os *Service Level Agreements* (SLA) negociados.

Em termos de interligação, o IPX opera sobre o modelo bilateral ou multilateral [11]. O modelo impulsionador do IPX foi o modelo multilateral, justificado pela necessidade da existência de um modelo de interligação IP global capaz de permitir o trânsito de tráfego IP de forma cómoda e simples.

Ao nível do serviço, o IPX é service aware, sendo que, ao contrário da Internet, este é capaz de identificar o serviço e realizar tratamentos distintos consoante os requisitos de QoS. Estes requisitos são cumpridos pelas partes intervenientes, respeitando os acordos negociados.

3. PROVA DE CONCEITO – VOIP *PEERING* INTERNACIONAL

As soluções VoIP *Peering* implementam na sua generalidade ambos os modelos (Bilateral e Multilateral), ficando ao critério do operador adotar o melhor que lhe convém. Em termos lógicos e, independentemente do modelo escolhido, torna-se necessário considerar funções de criação de rotas e funções de encaminhamento das mesmas. Na solução apresentada na figura seguinte, VoIP *peering* internacional (alvo de uma prova de conceito), são consideradas duas entidades para a realização destas funções, designadas por *Routing Function* e *Session Router*.

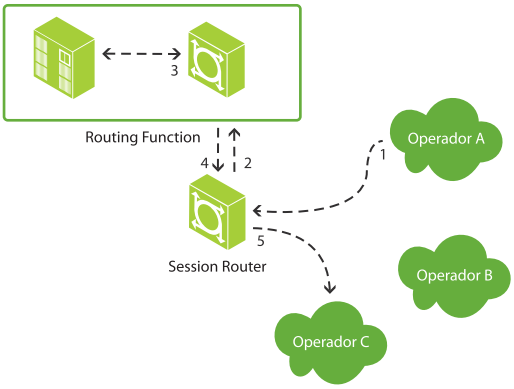


Figura 2. Arquitetura lógica de um modelo VoIP *Peering*

A entidade *Routing Function*, pode ser decomposta em dois componentes distintos: um dos componentes é responsável pela geração das tabelas de encaminhamento das sessões SIP, enquanto que o outro componente é responsável pela análise das tabelas e indicação das rotas a seguir.

O componente responsável pela geração das tabelas possui informação sobre quais os elementos de interligação de domínios que deve usar, para cada um dos destinos da tabela de encaminhamento de sessões SIP. O preenchimento da tabela obriga à escolha do elemento ou do conjunto de elementos que mais se adequa para um determinado destino. Esta escolha é realizada tendo em conta os critérios comerciais estabelecidos entre os operadores e a informação obtida dos *Interconnection Border Control Function* (I-BCFs) e dos *Media Gateway Controller Function* (MGCFs) para a caracterização das capacidades das ligações. Nesta fase são analisados os aspetos de QoS. Após a geração das tabelas, esta informação é passada para o outro componente, que através da análise dos dados da tabela e do conteúdo da sessão SIP, escolhe qual a melhor rota a utilizar.

Os componentes lógicos de geração e análise das tabelas podem ser implementados numa mesma entidade física, capaz de gerar e analisar as tabelas de modo a escolher as rotas.

A entidade *Session Router* é responsável por fazer pedidos ao *Routing Function* para obtenção de informações de encaminhamento e garante o encaminhamento das sessões SIP, sem tomar qualquer decisão na escolha das melhores rotas.

No âmbito da prova de conceito, foram considerados os elementos de *Routing Function* e *Session Router* implementados por três equipamentos distintos.

A função de *Session Router* é desempenhada pela entidade *Session Routing Proxy* (SRP) da Oracle (Acme Packet) [12], que se caracteriza por ser capaz de fazer a interligação entre I-BCFs, possuindo capacidade de encaminhamento de serviços *session-based* que podem ser serviços de voz, vídeo, serviços multimédia, entre outros. Em termos lógicos, o SRP apresenta-se como um elemento de processamento de pedidos, comunicando diretamente com os diferentes I-BCF e MGCFs, consoante a informação obtida do *Routing Function*. A nível funcional o SRP realiza *redirects* das sessões SIP para realizar os diferentes encaminhamentos.

O elemento *Routing Function* é implementado com recurso a duas entidades distintas, o OSPrey (do fabricante TransNexus) [13], responsável por funções de SIP *Routing Server* e por uma ferramenta de *Least Cost Routing* (LCR), responsável pela função de geração de tabelas de *routing* com base nos acordos comerciais, topologia das interligações e condições de rede.

O OSPrey server é um servidor de encaminhamento de sessões SIP [13], capaz de interagir com a ferramenta de LCR para obtenção da informação de encaminhamento necessária para responder aos pedidos recebidos pelo SRP. Para além de obter a informação da ferramenta de LCR, é também responsável pela obtenção de informação de *accounting* por parte dos I-BCFs e MGCFs, passando *Call Detail Records* (CDRs) para a ferramenta de LCR, com informação das ligações indicando tempos, qualidade de serviço entre outros. Desta forma, a

ferramenta de LCR tem toda a informação para que possa calcular dinamicamente as melhores rotas para os diferentes destinos.

A ferramenta de LCR assume a responsabilidade da geração das tabelas de encaminhamento e a indicação das rotas de encaminhamento internacional a nível da PSTN, tendo sido reutilizada para dar suporte de encaminhamentos internacionais a nível IP.

A figura seguinte apresenta a arquitetura de rede lógica usada na prova de conceito, em que existe um operador de trânsito que faz a interligação entre operadores IP e entre operador PSTN e operador IP.

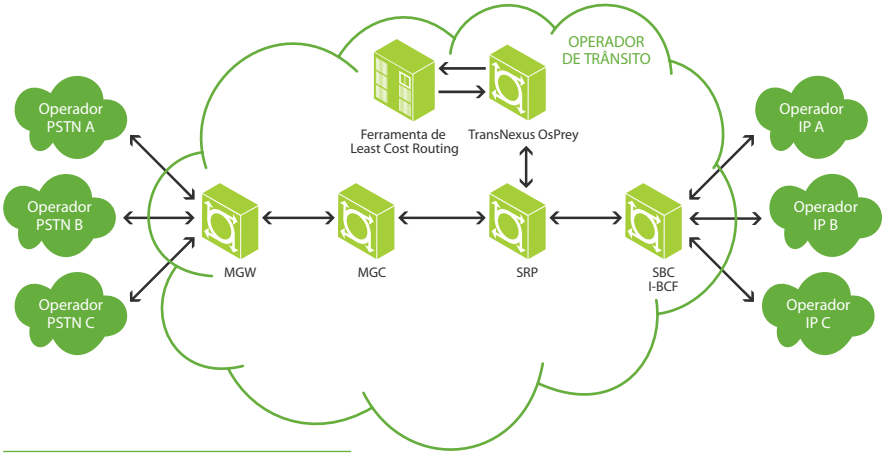


Figura 3. Diagrama lógico de prova de conceito

A figura seguinte apresenta o início de um fluxo de chamada num cenário de interligação de VoIP *peering* entre dois operadores IP.

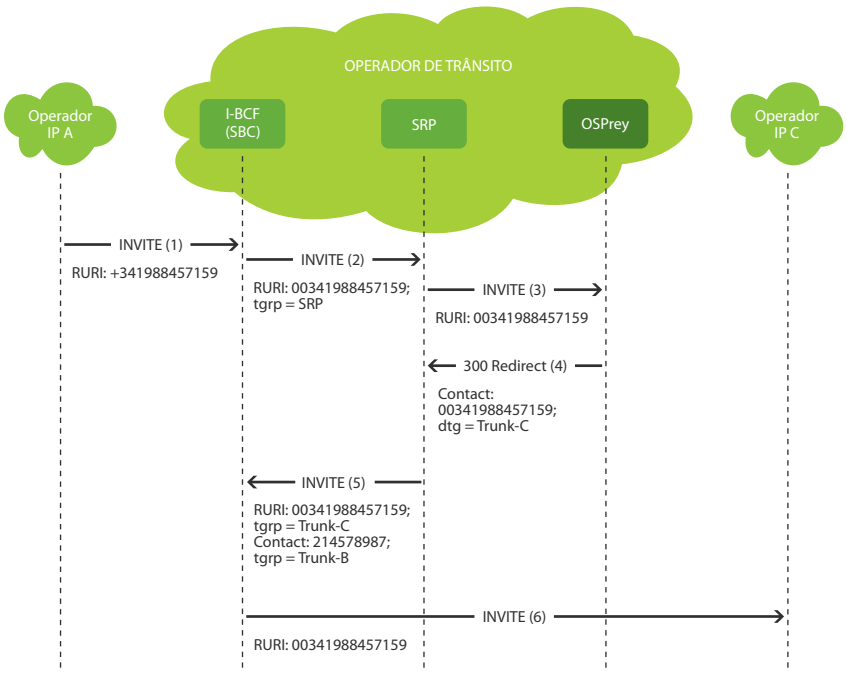


Figura 4. Diagrama de fluxo de início de sessão SIP entre dois Operadores IP

A mensagem SIP INVITE, originada no Operador A, entra na rede do operador de trânsito pelo I-BCF (SBC) e de seguida é encaminhada para o SRP, que por sua vez o encaminha para o OSPrey. Nesta fase, o OSPrey já detém toda a informação das tabelas de encaminhamento fornecidas pela ferramenta de LCR e envia uma mensagem SIP *Redirect* para o SRP, com a rota a usar para o operador C. De seguida, o SRP envia a mensagem SIP INVITE para o I-BCF que se encarrega de encaminhar a chamada para o Operador C.

Um outro exemplo de início de um fluxo de chamada num cenário de interligação de VoIP *peering* entre um Operador PSTN e um Operador IP, é apresentado na figura seguinte.

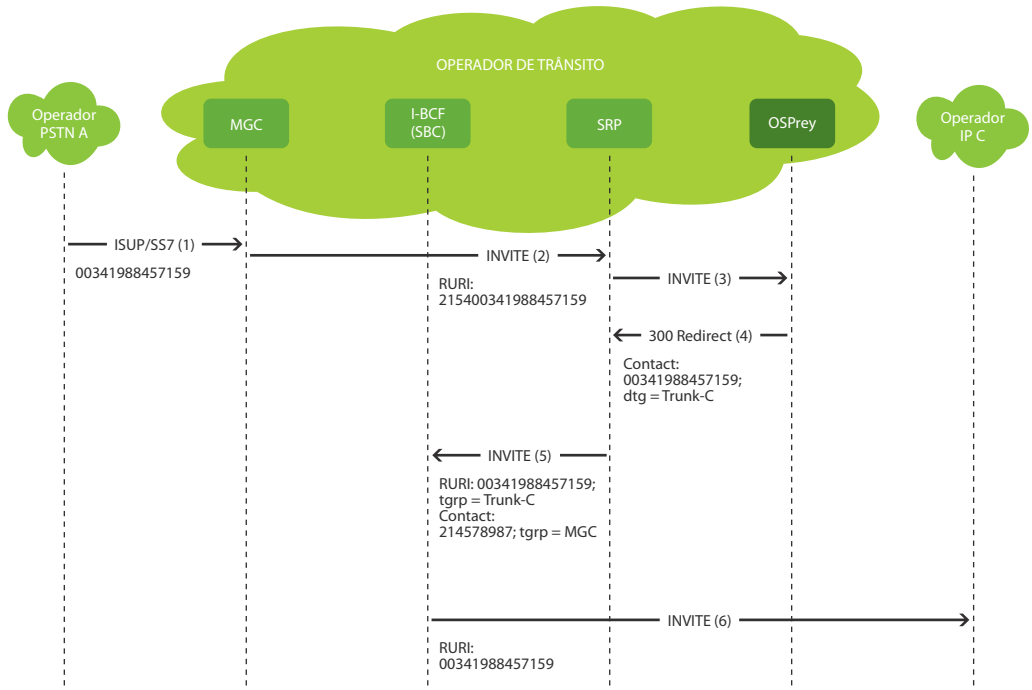


Figura 5. Diagrama de fluxo de início de sessão de um Operador PSTN para um Operador IP

A chamada originada no Operador PSTN entra na rede do operador de trânsito através da *Media Gateway Controller* (MGC), que adiciona um prefixo identificativo da rede de origem na mensagem SIP e encaminha-a para o SRP. À semelhança do fluxo anterior, o SRP encaminha a mensagem SIP para o OSPrey, que analisa a tabela de encaminhamento (obtida da ferramenta LCR) e passa a informação da rota a usar ao SRP, na mensagem SIP *Redirect*. De seguida, o SRP envia a mensagem SIP com a informação de encaminhamento para o I-BCF adequado, para que seja encaminhada para o operador C.

As questões de taxação, manutenção e qualidade de serviço são acauteladas com o tratamento da informação de *accounting*. A figura seguinte ilustra elementos envolvidos no tratamento desta informação.

Quando é terminada uma sessão SIP o I-BCF envia informação de *accounting* para o OSPrey. Através desta informação são gerados os CDRs e são posteriormente passados para a ferramenta de LCR. Desta forma é possível efetuar o ajuste das rotas consoante a ocupação das diferentes ligações, sendo considerados diversos factores tais como, a hora do dia, o tipo de sessões a decorrer, entre outras.

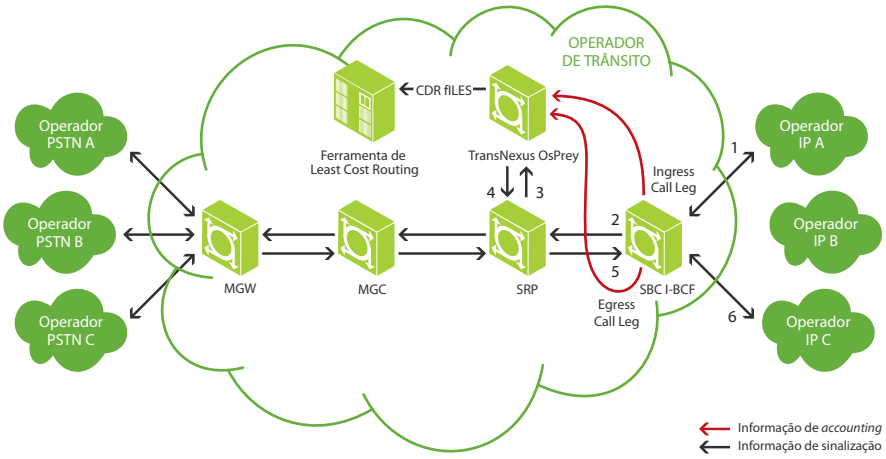


Figura 6. Diagrama lógico da transferência de informação de *accounting*

4. CONCLUSÕES

A utilização de soluções VoIP *Peering* representa uma evolução de serviços sobre a rede IP, permitindo aos operadores uma interligação do serviço de voz sobre a rede IP, com redução de custos face a soluções de interligação sobre a rede PSTN.

A prova de conceito descrita permitiu obter resultados bastantes interessantes, não só ao nível do sucesso da implementação efetuada, obtendo-se a melhor rota para cada caso, mas também ao nível do trabalho de integração efetuado por diferentes fornecedores, tendo-se conseguido operacionalizar uma solução que responde aos requisitos do operador.

As soluções de VoIP *Peering* apesar de terem vantagens, descritas neste artigo, face a soluções de voz sobre as redes legadas, ainda irão demorar algum tempo a serem massificadas, não por questões técnicas, mas sim por questões de interesse económico relacionadas com os operadores de PSTN que ainda dão este tipo de serviço. A organização das próprias soluções de VoIP *peering* e os requisitos regulatórios que existem e que ainda não se encontram totalmente esclarecidos podem também contribuir para um crescimento não tão rápido deste tipo de soluções.

REFERÊNCIAS

[1] 3GPP, *IP Multimedia Subsystem (IMS); Stage 2*, 3rd Generation Partnership (3GPP), 2013.

[2] IETF, "Session Peering for Multimedia Interconnect (SPEERMINT) Terminology," 2009.

[3] M. Lahti, *Analysis of VoIP interconnection evolution*, HELSINKI: HELSINKI UNIVERSITY OF TECHNOLOGY, 2008.

[4] P. T. Ltd, *VoIP Statistics – Market Analysis*; Q2 2012, 2012.

[5] J. Hodges, *Session Border Controllers: Addressing Tomorrow's Requirements*, Heavy Reading, 2011.

[6] "VoIP Peering & the Future of telecom Network Interconnection," *Heavy Reading*, 8 2006.

[7] ITU-T, "GSR Discussion Paper," 2009.

[8] TransNexus, "Optimize Your VoIP Network By Managing Routing Decisions," 2012. [Online]. Available: http://www.transnexus.com/index.php/what-is-least-cost-routing/least-cost-routing.

[9] XConnect, "Bridging the VoIP Islands," 2009.

[10] GSM Association, *Guidelines for IPX Provider networks (Previously InterService Provider IP Backbone Guidelines)*, 2013.

[11] G. Association's, "IP eXchange; Providing a quality based solution," 2012.

[12] A. Packet, "Net-Net Session Router," 2012.

[13] TransNexus, "OSPrey-32," TransNexus, 2012. [Online]. Available: http://www.transnexus.com/index.php/osprey32.

CVS DOS AUTORES

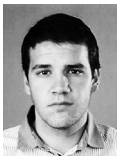
António Manuel Amaral, concluiu em 2001, a licenciatura em Engenharia Electrónica e de Telecomunicações, pela Universidade de Aveiro. Ingressou nesse ano no Instituto de Telecomunicações de Aveiro como investigador na área de Redes. Concluiu em 2006, o Mestrado em Engenharia Electrónica e de Telecomunicações, pela Universidade de Aveiro, defendendo a dissertação de mestrado intitulada de "Encaminhamento Multicast em Redes IP". Ingressou na PT Inovação em 2006 e está atualmente incorporado na direção de Instalação, Entrega e Suporte de Plataformas de Serviços, na divisão de Especificação, Projecto e Integração de Rede, desempenhando do papel de Team-Leader de uma equipa responsável pelo desenvolvimento de Soluções e Serviços Convergentes em redes IMS.

David Gonçalves, concluiu a licenciatura em Engenharia Informática, pelo Instituto Politécnico do Porto em 2011, iniciando no mesmo ano a pós-graduação na Universidade do Minho de Engenharia de Redes e Serviços de Comunicações. Em 2013 ingressou na PT Inovação na área de Redes de Próxima Geração, tendo realizado o projeto de mestrado na área das redes IMS. Atualmente encontra-se na mesma área estando a desenvolver trabalhos sobre os equipamentos de entrada de rede.

José Carlos Silva, concluiu o Bacharelato em Engenharia Informática, pelo Instituto Politécnico de Bragança em 2002, e licenciou-se em Engenharia de Sistemas e Informática, pela Universidade do Minho em 2005. Ingressou nesse mesmo ano a PT Inovação, estando desde então ligado à área de Redes de Próxima Geração. Atualmente integra o departamento de Infraestrutura de Plataformas de Serviços, onde desempenha funções de desenho e integração de soluções IMS.

Francisco Fontes, licenciado em Engenharia Electrotécnica e de Computadores, ramo de Electrónica e Telecomunicações, pelo Instituto Superior Técnico (Setembro de 1991) e Doutorado pela Universidade Politécnica de Madrid (Novembro de 2000) na área de gestão distribuída de redes de telecomunicações. Em Setembro de 1991 iniciou a sua actividade profissional na PT Inovação (CET) tendo-se especializado em tecnologias de rede de banda larga. De 2002 a 2012 foi Professor Auxiliar Convidado da Universidade de Aveiro/DETI, para as áreas de redes IP aplicadas às telecomunicações. Actualmente os seus interesses situam-se na área das arquitecturas de redes, em especial na sua evolução para arquitecturas RPG All-IP, com ênfase no IMS. É especialista em tecnologias de rede local e acesso, IPv6 e Multicast.

13 Automatização de Testes SIP em Redes IMS


ANTÓNIO AMARAL DAVID GONÇALVES JOSÉ SILVA

PALAVRAS CHAVE
Automatização de testes, IMS, SIP

A qualidade de um projeto encontra-se diretamente relacionada com a capacidade dos testes abrangerem todos os requisitos existentes, garantindo-se que quanto mais testes forem efetuados, menos problemas serão reportados.

A lista de requisitos de um projeto pode ser densa e complexa, havendo a necessidade de os decompor em requisitos mais simples para os poder mapear em testes específicos. O tempo e os recursos gastos para a execução dos testes, dependem claramente do método usado para a implementação deste processo. A automatização do processo de testes apresenta-se como um conceito chave, já que permite melhorar o tempo de execução dos testes e minimizar os erros de análise dos resultados obtidos.

Os testes para validação de novas versões de *firmware* são fundamentais para garantir que todas as funcionalidades (novas e antigas) são garantidas e que os problemas são corrigidos, se forem automatizados resultam num ganho de tempo significativo, face à validação dos mesmos usando um processo manual.

No entanto, embora as vantagens adjacentes à utilização de testes automáticos sejam evidentes, o processo de mapeamento dos requisitos para testes é complexo, porque a informação simples de analisar por um humano por vezes tem de ser decomposta em elementos com maior detalhe que possam ser analisados por máquinas.

Baseado na experiencia prática de validação do *firmware* dos *Session Border Controllers* (SBCs) [1] de soluções implementadas em rede de operador, este artigo apresenta uma metodologia de desenvolvimento capaz de, através da análise dos requisitos, mapeá-los em testes automáticos, diminuindo desta forma o esforço contínuo que é necessário para realizar entregas recorrentes aos clientes.

1. INTRODUÇÃO

As redes de telecomunicações encontram-se em constante evolução, impulsionadas pela constante procura de soluções cada vez mais ricas e convergentes que por sua vez têm vindo a contribuir para a mudança de paradigma no desenvolvimento dos próprios serviços, bem como para a mudança para uma arquitetura de rede *IP Multimedia Subsystem* (IMS) [2].

A arquitetura IMS, caracteriza-se por ser uma arquitetura funcional estratificada e que tem por princípio a disponibilização de serviços do tipo *session-based* a terminais servidos por redes de acesso distintas [3]. Esta estratificação permite a criação de um sistema interoperável, capaz de disponibilizar um conjunto vasto de serviços, à custa da garantia dos requisitos de qualidade de serviço e considerando sempre as questões de mobilidade. A arquitetura apresenta-se decomposta em três níveis: transporte, controlo, serviço.

Ao nível do transporte, são endereçadas todas as questões de conectividades com as diferentes redes de acesso, permitindo a sua dissociação com os serviços disponibilizados [4]. Ao nível do controlo, trata toda a sinalização associada ao estabelecimento de sessões e são consideradas as questões relacionadas com a autenticação e taxação. Finalmente ao nível do serviço, encontra-se a disponibilização de serviços avançados sobre a rede IMS, sendo este nível responsável pela disponibilização de serviços de redes legadas (serviços disponibilizados pelo CAMEL [5] nas redes de segunda-geração) ou a disponibilização de serviços inovadores desenvolvidos de raiz para as redes IMS.

Caracterizando-se como sendo uma arquitetura modular capaz de disponibilizar um conjunto extenso de serviços, as redes IMS apresentam-se como o passo natural na evolução das redes dos operadores e já o começam a ser atualmente, permitindo-lhes desenvolver soluções com requisitos e funcionalidades distintas, indo sempre de encontro às necessidades dos utilizadores.

A necessidade de validar os requisitos na sua totalidade apresenta-se como um fator fundamental para a determinação do sucesso ou insucesso de uma solução. Este princípio, está associado a qualquer tipo de produto/serviço, onde se enquadram todas as soluções em redes IMS. Contudo, a realização de validações da totalidade dos requisitos de forma manual caracteriza-se por ser processo moroso, de elevado consumo de recursos humanos e suscetível a erros. A automatização das validações apresenta-se, assim, como o caminho indicado para a obtenção de uma solução robusta e capaz de ter em consideração todos os requisitos.

Sendo o protocolo *Session Initiation Protocol* (SIP) o protocolo de sinalização usado em redes IMS e sendo este protocolo composto por diferentes tipos de mensagens, com diferentes *headers*, é fundamental executar processos de validação da integração dos componentes que implementam este protocolo na rede IMS. Este ponto assume particular importância quando se está perante elementos de rede responsáveis por adaptações protocolares de *headers* SIP, como é o caso do SBC, que é o elemento de entrada SIP na rede IMS do operador, fazendo a interligação entre a rede de acesso - onde está o cliente - e a rede do operador.

2. MOTIVAÇÃO

Os testes a realizar para validar uma solução estão diretamente relacionados com o que se pretende

>>>

validar e com a maturidade da solução implementada. Desta forma, pode-se separar o âmbito dos testes em testes funcionais, testes de sanidade, testes de aceitação e testes de regressão.

Os testes funcionais têm como objetivo validar o comportamento funcional da solução ou componente em causa. Os testes de sanidade pretendem validar se a solução apresenta um nível de maturidade para ser submetida a testes mais detalhados com um nível de profundidade maior. Os testes de aceitação pretendem validar todos os casos num ambiente o mais próximo do ambiente de produção e, finalmente, os testes de regressão são executados em ambiente de produção, de forma a confirmar que a alteração introduzida na solução (quer seja por atualização de *hardware,* quer seja por atualização de *software*) não tem efeitos colaterais que possam comprometer as funcionalidades existentes.

Como se verifica, existe um conjunto extenso de pontos de validação de requisitos, sendo que, para cada um deles, dever-se-á realizar os testes que mapeiam os requisitos na sua totalidade ou um subconjunto destes. O que se pretende com a solução de automatização de testes SIP em redes IMS é definir procedimentos para a validação das soluções suportadas num determinado equipamento, sendo neste caso particular consideradas as soluções suportadas no SBC.

Na operacionalização da solução automática de testes SIP foi necessário isolar o SBC da rede e ter o controlo dos *inputs* e *outputs* que condicionam o seu comportamento. Posteriormente, definiu-se um conjunto de requisitos sujeitos a validação automática e foi necessário desenvolver mecanismos capazes de executar os testes e realizar capturas ao nível SIP, para posteriormente se fazer a análise dos resultados finais. Tendo controlo sobre os *inputs* e *outputs*, torna-se possível simular o ambiente real aplicado apenas ao SBC, já que se consegue replicar os fluxos de sessões SIP iguais aos que se tem na rede do operador, à custa de ferramentas de emulação de tráfego responsáveis por executar cenários de validação. Neste caso particular, foi usado o SIPp [6], que permite a geração de tráfego SIP. Esta ferramenta caracteriza-se por apresentar um bom nível de maturidade ao nível SIP e por estar muito bem documentada com exemplos que caracterizam diferentes fluxos de chamadas SIP, o que simplificou a sua utilização.

A operacionalização dos fluxos de testes carregados na ferramenta SIPp, se for feita de forma manual, torna-se num processo lento e sujeito a erros. Por esse motivo, utilizou-se um processo automático para que, com base em capturas de redes reais, fossem operacionalizados os casos de testes para os cenários pretendidos.

No processo de execução dos testes foi necessário ter em consideração que os cenários de envio e receção devem ser executados de forma simultânea e devem ter a capacidade de registar as evidências respetivas.

3. DESENHO DE SOLUÇÃO

A complexidade do processo de automatização de testes não permite que a validação seja realizada através de uma aplicação unitária que atua de forma completamente independente do utilizador. Através de procedimentos normalizados, consegue-se automatizar a grande maioria dos processos de validação com resultados bastante satisfatórios quando comparados com o processo de validação manual existente.

OPERACIONALIZAÇÃO DOS CENÁRIOS

Para a operacionalização dos cenários é seguido o fluxo descrito na Figura 1, sendo considerados os seguintes aspetos:

- Configurações de rede
- Verificação das configurações de serviço
- Implementação das configurações

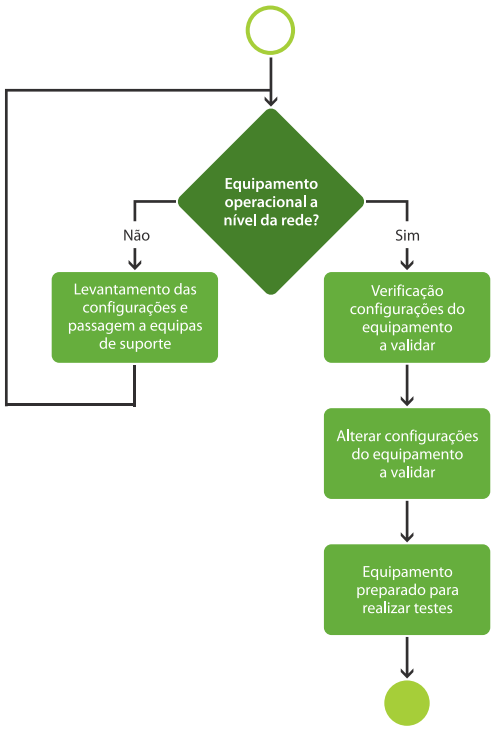


Figura 1. Fluxo de operacionalização dos cenários

Configurações de rede

De forma a ter um cenário igual ao do ambiente de produção, operacionalizou-se toda a rede envolvente para replicar o cenário real onde o SBC opera, garantindo-se que todo o endereçamento do IP (*Internet Protocol*) e das VLANs (*Virtual Local Access Networks*) é igual ao do ambiente de produção.

A grande vantagem desta abordagem passa por facilitar a importação das configurações e execução dos testes sobre as condições implementadas na rede de produção, sem haver necessidade de envolver os próprios SBCs que estão a dar o serviço real.

Verificação das configurações

Antes da realização das alterações nas configurações do equipamento é necessário fazer o levantamento das configurações que necessitam de ser modificadas. Estas verificações encontram-se fortemente associadas com as configurações do equipamento, sendo necessário possuir conhecimento razoável do equipamento para realizar o levantamento das configurações que necessitam de sofrer alterações de modo a disponibilizarem serviço. No caso específico de validação de soluções no SBC, foi necessário separar funcionalidades tais como desativação dos cenários de registo quando se realizam testes de estabelecimento de sessão, desativação do envio das mensagens *keep alive* para não ter testes com entropia, entre outros. Para facilitar a validação, desenvolveu-se uma aplicação simples capaz de realizar de forma rápida o levantamento das configurações a alterar.

Implementação das configurações

Após ter o equipamento preparado na rede e ter sido feito o levantamento das configurações, é necessário proceder à sua implementação. Para tal, são usados *scripts* aplicados a clientes SSH (*Secure Shell*) e ou Telnet com as configurações a atualizar. A ferramenta de SSH/Telnet escolhida para a realização das configurações foi o ExtraPutty [7], que se liga ao equipamento e executa um conjunto de comandos com as alterações requeridas. Para cada cenário a validar é necessário um *script* de arranque (*script* associado ao ExtraPutty) distinto, de modo a habilitar as configurações do equipamento.

CRIAÇÃO DE TESTES

No processo de criação de testes foram consideradas duas fases: a criação de *scripts* SIPp e a aplicação de expressões regulares.

Criação dos *scripts* SIPp

Para a criação dos *scripts* SIPp implementou-se um processo automático que usa a ferramenta *open source* sniff2sipp [8] para gerar os *scripts*, tendo como *input* as capturas de rede no formato ".pcap". Contudo, para a ferramenta operar devidamente, foi necessário adaptar a análise da pilha protocolar usada pela ferramenta, dando a possibilidade de existirem cabeçalhos com identificação de VLANs e tornando, assim, possível analisar a totalidade das capturas realizadas em ambiente de produção. Para além da alteração da pilha protocolar suportada, foi dada a possibilidade de inserir valores dinâmicos nos *scripts* SIPp através da definição de variáveis.

Aplicação de expressões regulares

A criação dos cenários, por si só, não permite a realização da totalidade das validações, já que têm que ser feitas com um nível de detalhe aprofundado, consequência da necessária validação de todo o conteúdo das mensagens SIP. Usando a possibilidade do SIPp suportar expressões regulares, definiu-se um conjunto de expressões regulares aplicáveis a cada cenário.

A utilização das expressões regulares para validação dos conteúdos dos cabeçalhos das mensagens SIP possibilita a realização de um vasto conjunto de análises, já que podem ser considerados vários parâmetros. De seguida, apresenta-se um exemplo de uma expressão regular aplicada a uma mensagem SIP INVITE, onde é analisado o conteúdo do *header* From, comparando o valor do *header* com o valor recebido na mensagem anterior (Figura 2).

```
<recv request="INVITE" crlf="true">
 <action>
   <ereg regexp=".*" search_in="hdr" header="From:" assign_to="1"
   </ereg>

   <log message="From is [last_From]. Custom header is [$1]"/>
  </log>
 </action>
</recv>
```

Figura 2. Exemplo de uma expressão regular em SIPp

>>>

Em todos os testes é necessário indicar a expressão regular que se pretende validar, indicando o valor ou o conjunto de valores que seriam expectáveis receber, que excerto da mensagem se pretende analisar, se é pretendido analisar a mensagem na sua totalidade "msg" ou se é apenas sobre um cabeçalho específico "hmr" e se é pretendido que se armazenem em variáveis os valores pretendidos.

Se o resultado das expressões regulares for positivo, então o teste é dado como passado. Caso contrário, o teste é dado como falhado e é indicado qual o motivo para o insucesso do mesmo.

EXECUÇÃO DE TESTES

O procedimento de execução de testes engloba todas as atividades associadas com a execução, verificação dos comportamentos, forma de realizar os testes e garantia de registo de evidências.

Na execução dos testes foram considerados os atores apresentados na figura seguinte, cada um deles implementado em máquinas distintas.

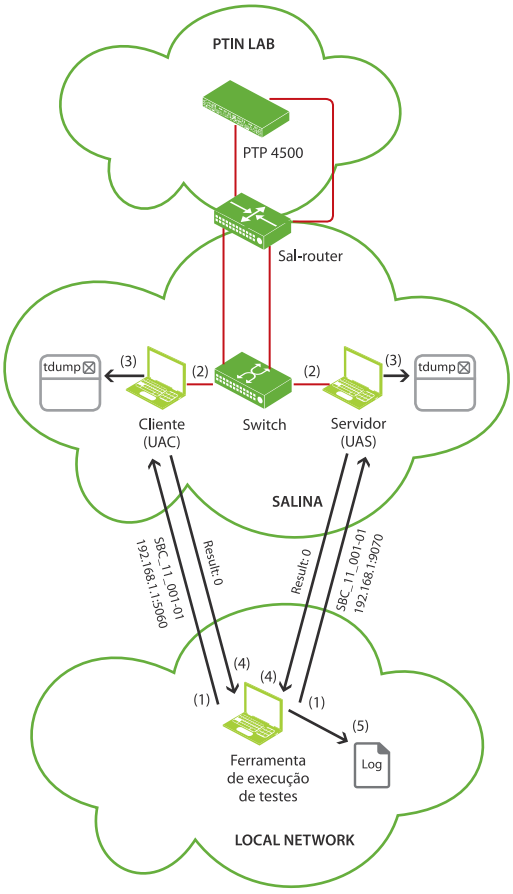


Figura 3. Diagrama Lógico de execução de testes

A ferramenta de execução de testes é o ator responsável por iniciar a execução da bateria de testes, indicando tanto ao cliente como ao servidor quais os *scripts* SIPp que devem ser executados. Os testes a validar são executados de forma síncrona entre o elemento responsável pela geração do *input*, fazendo este o papel de *User Agent Client* (UAC), e o elemento que recebe o *output* gerado pelo equipamento a validar, sendo este denominado de *User Agent Server* (UAS). De forma a realizar a validação, pode-se optar por duas soluções distintas para a implementação da ferramenta de execução de testes. Por um lado, pode-se desenvolver uma aplicação de controlo de máquinas sobre ligações remotas para a execução dos testes. Por outro lado, podem-se integrar os testes numa plataforma de testes automáticos (ex. Cucumber) capaz de executar um conjunto de testes controlando a plataforma e as máquinas UAC e UAS. No caso específico da validação dos SBCs, optou-se pelo desenvolvimento de raiz de uma aplicação de controlo das máquinas, não tendo sido possível fazer a integração das validações numa plataforma de validação automática.

O registo de evidências pode ser obtido de formas distintas dependendo da plataforma escolhida para a realização da validação de testes. Através da plataforma Cucumber, possui-se acesso a um conjunto de *reports* detalhados sobre o funcionamento do teste. Ao nível aplicacional obtém-se evidências através da integração de comandos *tcpdump* na execução dos *scripts* SIPp, bem como através de *reports* globais da bateria de testes executados.

O *report* global de execução da bateria de testes apresenta-se na Figura 4, onde se mostra informação sobre a bateria de testes, estatísticas sobre as validações e informação detalhada da execução de cada teste individual.

```
Registo de execução de testes
------------------------------
Versão:
Data:

Resumo
------
1) Testes com sucesso    :
2) Testes com insucesso  :
2.1) Expressão regulares :
2.2) Questões de Rede    :

Detalhes
--------
```

Figura 4. Relatório de execução de testes

A informação da bateria de testes indica o estado dos testes e a data da execução. A informação estatística apresenta os resultados, de forma agregada, da execução da bateria de testes, a percentagem de testes com sucesso e a percentagem de testes com insucesso. A informação detalhada indica para cada teste o *output* obtido pela ferramenta de validação associado à execução do SIPp, tanto do lado do cliente como do lado do servidor.

Nas evidências obtidas pelo *tcpdump* é possível analisar a informação do fluxo associado ao *User Agent* (UA). A Figura 5 apresenta um exemplo da informação capturada.

```
192.168.7.13 -> 192.168.8.15 SIP 668 Request: REGISTER sip: dom.ims.pt    (fetch bindings)
192.168.8.15 -> 192.168.7.13 SIP 566 Status: 401 Unauthorized   (0 bindings)
192.168.7.13 -> 192.168.8.15 SIP 921 Request: REGISTER sip: dom.ims.pt    (fetch bindings)
192.168.8.15 -> 192.168.7.13 SIP 691 Status: 200 OK    (0 bindings)

192.168.7.13 -> 192.168.8.15 SIP/SDP 949 Request: INVITE tel:123456789, with session description
192.168.8.15 -> 192.168.7.13 SIP 330 Status: 100 Trying
192.168.8.15 -> 192.168.7.13 SIP 481 Status: 180 Ringing
192.168.8.15 -> 192.168.7.13 SIP/SDP 965 Status: 200 OK, with session description
192.168.7.13 -> 192.168.8.15 SIP 495 Request: ACK tel:123456789
192.168.7.13 -> 192.168.8.15 SIP 572 Request: BYE tel:123456789
192.168.8.15 -> 192.168.7.13 SIP 363 Status: 200 OK

192.168.7.13 -> 192.168.8.15 SIP 665 Request: REGISTER sip: dom.ims.pt    (fetch bindings)
192.168.8.15 -> 192.168.7.13 SIP 566 Status: 401 Unauthorized    (0 bindings)
192.168.7.13 -> 192.168.8.15 SIP 918 Request: REGISTER sip: dom.ims.pt    (fetch bindings)
192.168.8.15 -> 192.168.7.13 SIP 618 Status: 200 OK   (0 bindings)
```

Figura 5. Informação capturada na execução de testes automáticos

A captura apresenta apenas os fluxos existentes entre o SBC e a respetiva máquina UA onde a captura é realizada. Em relação ao fluxo da captura, só é constituído por mensagens associadas ao teste que se pretende validar, filtrando os restantes pacotes. Para cada teste são gerados dois ficheiros ".pcap", um com o comportamento de *input* do SBC e outro com o *output*.

4. TESTES E RESULTADOS

Os resultados obtidos com o processo automático permitiram concluir um ganho efetivo de tempo no processo de execução dos testes de validação de *firmware* do SBC. Estes resultados assumem particular relevância quando se trata de um elemento de rede SIP responsável por um elevado conjunto de manipulações de diferentes *headers* em diferentes mensagens SIP e cuja análise do resultado dessas manipulações é extremamente morosa e sujeita a erros.

Num processo manual, para a validação do resultado dos testes é necessário configurar terminais, realizar o teste capturando evidências, filtrar as evidências para o teste em questão, fazer a análise do fluxo e de cabeçalhos específicos, descrever o resultado do teste, armazenar as evidências e analisar os resultados finais. Estes passos resultam em largos minutos desde que se inicia o processo de execução, até que se termina o teste com a respetiva análise. Conjuntos de testes anteriores para validação de versões de *firmware* usando o processo manual registam tempos médios de execução de um teste na ordem dos 60 minutos.

Nos mesmos testes de validação de *firmware*, mas usando o processo automático, registam-se tempos médios de execução de um teste na ordem dos 10 segundos, o que se traduz claramente num ganho bastante significativo de execução do teste. Neste período de tempo são executados os cenários com os números de terminal associados, são validados os fluxos (através do controlo restrito do SIPp sobre as mensagens que são esperadas receber ou não), são validados cabeçalhos específicos (através das expressões regulares), são capturadas evidências, lançados relatórios de execução e é registado o sucesso ou insucesso do teste.

O gráfico presente na Figura 6 apresenta alguns resultados obtidos para o processo manual e para o processo automático. No processo manual, é apresentado o resultado da execução de 25 testes realizados por uma pessoa com larga experiência em SBCs e em redes IMS, conhecendo em detalhe o protocolo SIP e os requisitos associados a cada teste. No processo automático, são apresentados dois casos de execução de 300 testes, um deles executados por quem desenvolveu e implementou o processo de automatização

e outro caso realizado por parte de uma pessoa que não participou no desenvolvimento do processo, ambos na sua fase de iniciação ao SBC e às redes IMS.

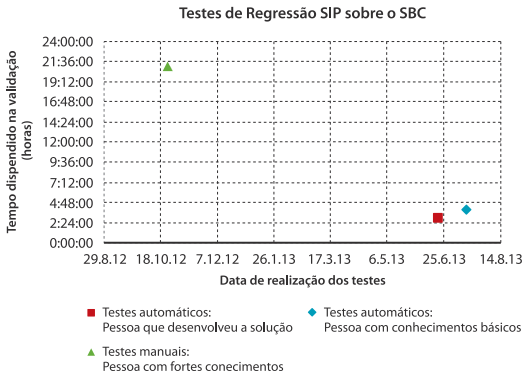


Figura 6. Comparação dos tempos de validação de modo manual e automático

Analisando os resultados anteriores constata-se que, por um lado, com o processo automático consegue-se executar um maior número de testes, com uma redução significativa do tempo total de execução dos testes quando comparado com o tempo total de execução de um menor número de testes através do processo manual. Por outro lado, com o processo automático, a pessoa responsável pela execução dos testes não tem que ter um conhecimento aprofundado da tecnologia em questão, nem dos requisitos de cada testes, podendo esta tarefa de execução ser entregue às equipas responsáveis para execução dos testes de outras plataformas.

5. CONCLUSÕES

Ao ritmo acelerado a que as redes de telecomunicações evoluem, é fundamental diminuir tempos associados à realização de tarefas comuns, enquadrando-se nestas tarefas a validação de novas versões de *firmware* dos SBCs em redes IMS.

A consolidação do processo automático da validação de *firmware* do SBC requereu um esforço inicial significativo de forma a operacionalizar todo o processo. No entanto, este esforço inicial foi claramente compensado com o resultado final, já que com o processo implementado consegue-se, por um lado, executar um conjunto de testes mais alargado num período de tempo muito inferior e, por outro lado, é possível aumentar o número de validações efetuadas às mensagens SIP, à custa de uma diminuição de falhas de análise dos resultados obtidos, já que são usados comportamentos determinísticos no processo de análise.

O processo e métodos descritos para a validação automática SIP de versões de *firmware* ao nível do SBC podem ser aplicáveis a outros componentes das redes IMS que implementem a interface protocolar SIP, tais como o *x-Call Session Control Function* (x-CSCF), *Application Server* (AS), entre outros.

Finalmente importa referir que a solução de automatização de testes, para além do SIP, pode ser aplicada a outros protocolos (por exemplo o protocolo *diameter*), aplicando à solução geradores de tráfego que implementem esse protocolo.

REFERÊNCIAS

[1] J. Hodges, *Session Border Controllers: Addressing Tomorrow's Requirements*, Heavy Reading, 2011.

[2] 3GPP, *IP Multimedia Subsystem (IMS); Stage 2*, 3rd Generation Partnership (3GPP), 2013.

[3] G. Camarillo e M. A. Garcia-Martin, The 3G IP multimedia subsystem IMS – merging the internet and the cellular worlds (2. ed.)., Wiley, 2006.

[4] Ericsson, *IMS – IP Multimedia Subsystem, The value of using the IMS architecture*, 2004.

[5] 3GPP, *Customized Applications for Mobile network Enhanced Logic (CAMEL) Phase X; Stage 2*, 3rd Generation Partnership (3GPP), 2010.

[6] R. Gayraud, "SIPp," 2013. [Online]. Available: http://sipp.sourceforge.net/.

[7] ExtraPuTTY, "ExtraPuTTY," 2013. [Online]. Available: http://www.extraputty.com/.

[8] T. Wilson, *sniff2sipp*, Digium, 2008.

CVS DOS AUTORES

António Manuel Amaral, concluiu em 2001, a licenciatura em Engenharia Electrónica e de Telecomunicações, pela Universidade de Aveiro. Ingressou nesse ano no Instituto de Telecomunicações de Aveiro como investigador na área de Redes. Concluiu em 2006, o Mestrado em Engenharia Electrónica e de Telecomunicações, pela Universidade de Aveiro, defendendo a dissertação de mestrado intitulada de "Encaminhamento Multicast em Redes IP". Ingressou na PT Inovação em 2006 e está atualmente incorporado na direção de Instalação, Entrega e Suporte de Plataformas de Serviços, na divisão de Especificação, Projecto e Integração de Rede, desempenhando do papel de Team-Leader de uma equipa responsável pelo desenvolvimento de Soluções e Serviços Convergentes em redes IMS.

David Gonçalves, concluiu a licenciatura em Engenharia Informática, pelo Instituto Politécnico do Porto em 2011, iniciando no mesmo ano a pós-graduação na Universidade do Minho de Engenharia de Redes e Serviços de Comunicações. Em 2013 ingressou na PT Inovação na área de Redes de Próxima Geração, tendo realizado o projeto de mestrado na área das redes IMS. Atualmente encontra-se na mesma área estando a desenvolver trabalhos sobre os equipamentos de entrada de rede.

José Carlos Silva, concluiu o Bacharelato em Engenharia Informática, pelo Instituto Politécnico de Bragança em 2002, e licenciou-se em Engenharia de Sistemas e Informática, pela Universidade do Minho em 2005. Ingressou nesse mesmo ano a PT Inovação, estando desde então ligado à área de Redes de Próxima Geração. Atualmente integra o departamento de Infraestrutura de Plataformas de Serviços, onde desempenha funções de desenho e integração de soluções IMS.

14 PCRF com Controlo de Congestão


CARLOS MARQUES


CARLOS PARADA


CARLOS RODRIGUES


FILIPE RODRIGUES


FRANCISCO FONTES


MIGUEL SANTOS


PEDRO NEVES


SUSANA SARGENTO


TELMA MOTA


TIAGO CARDOSO

PALAVRAS CHAVE
3GPP, LTE, PCC, PCRF, PCEF, TDF, CAC, RAN, eNB, EPC, Wi-Fi, ip-Raft

Na atualidade, o controlo de recursos em redes móveis (mas não só) é feito através do componente PCRF (*Policy and Charging Rules Function*), definido pelo 3GPP, o qual constitui a base do produto ip--Raft da PT Inovação. Este controlo é feito com base em alguma informação que advém dos elementos de rede, como sejam o tráfego cursado por um cliente, o seu agregado diário ou mensal, ou a rede de acesso na qual está ligado (3G, 4G, Wi-Fi, etc.). No entanto, esta informação não é suficiente para perceber se alguns dos elementos da rede se encontram congestionados, não sendo, consequentemente, capaz de lidar com a congestão da forma mais eficaz. Por isso, poderá impor penalizações a clientes de maior valor ou a serviços mais sensíveis, em vez de selecionar aqueles que sofreriam um menor impacto. Igualmente importante, não consegue evitar que a congestão aconteça, o qual poderia ser conseguido através de uma atuação preventiva, por exemplo, balanceando ou migrando os terminais para outras tecnologias, sempre que estas se encontrem disponíveis.

Assim, o objetivo do PCRF CAC (Call Admission Control) é o de enriquecer o PCRF (produto ip-Raft) já existente, com a capacidade de lidar de forma mais eficaz com episódios de congestão, levando-o a tomar a melhor opção, de acordo com as regras definidas pelo operador. Para isso, o PCRF recolhe informação a partir de variadas fontes, de forma a ter uma visão global do estado da rede, podendo assim aplicar políticas de forma mais competente. Com esta capacidade, o ip-Raft passará a estar na linha da frente do que melhor se faz nesta área, posicionando-se assim como um produto altamente inovador e capaz de ombrear com a concorrência.

Este projeto foi realizado no âmbito de um Projeto Inovação (PI) "PCRF" 2012/2013, em parceria com o Instituto de Telecomunicações (IT) de Aveiro. Como resultado deste projeto, foi realizada uma implementação do modelo e criado um protótipo, o qual permitiu efetuar a prova de conceito. Este artigo mostra os resultados obtidos, sendo que os desenvolvimentos concretizados deverão passar a fazer parte do produto ip-Raft já a partir da sua versão 6.2 ou 6.3.

1. INTRODUÇÃO

A crescente utilização de smartphones e tablets faz com que os clientes consumam cada vez mais serviços, nomeadamente vídeo, e gerem cada vez mais tráfego nas redes móveis. Apesar do crescente investimento dos operadores, o aumento da capacidade das redes não conseguirá acompanhar completamente o incremento do consumo, especialmente numa altura em que muitos outros *players* (OTTs) têm obtido receitas crescentes na cadeia de valor que antes revertia completamente para o lado dos operadores. Assim, será inevitável que algum tipo de congestão (*bottleneck*) possa afetar a rede, quer esta seja ocasional e temporária, quer tenha um caráter mais permanente.

Hoje em dia, para resolver a congestão, os operadores investem em infraestruturas com maior capacidade e priorizam os subscritores/serviços. Mas estas medidas por si só não resolvem o problema e continuam a existir momentos (picos) de congestão. Para enfrentar realmente problema questão do congestionamento, o operador necessita de saber as causas do problema, onde ocorrem, e como aplicar ações preventivas ou corretivas em tempo real. Estas ações podem ser aplicadas a células congestionadas, a determinados tipos de serviços ou aplicações, a clientes, entre outros, conforme seja mais apropriado.

O 3GPP define a arquitetura PCC (*Policy and Charging Control*), a qual tem como principal função o controlo dos recursos na rede para garantir a QoS dos subscritores. A arquitetura PCC (a cinzento na Figura 1) interage com a arquitetura LTE como se mostra. O PCRF (*Policy Rules and Charging Function*) é o elemento central nesta arquitetura, sendo aquele que controla os recursos alocados a cada subscritor (tendo em conta as reservas necessárias, chamada de voz, vídeo, entre outros), ou que faz o policiamento (*throttling*) do tráfego quando o subscritor vai para além dos limites de consumo estabelecidos para o seu perfil. No limite, pode mesmo barrar o tráfego, não permitindo qualquer serviço ao subscritor.

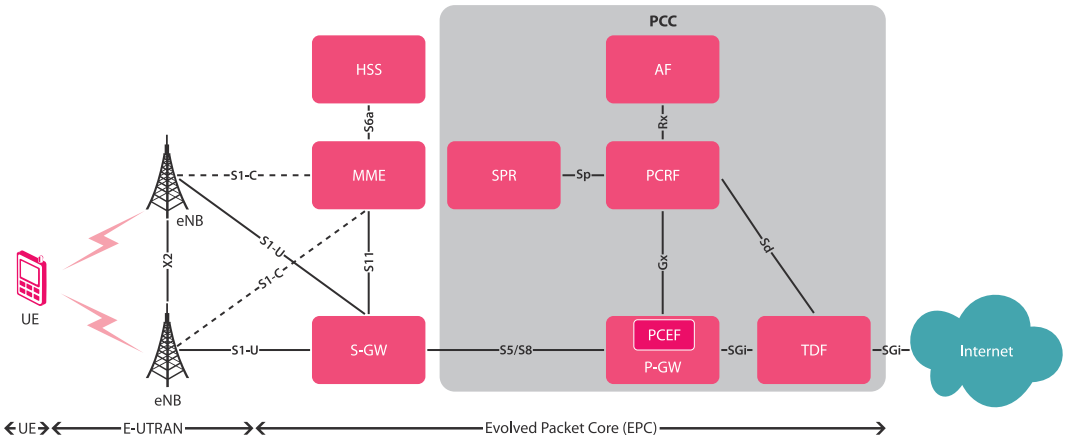


Figura 1. Arquitetura 3GPP LTE e PCC

Tradicionalmente, o PCRF tem em conta a informação que lhe é enviada pelo P-GW/PCEF sobre o tráfego cursado pelos subscritores, assim como informação sobre o perfil dos mesmos (SPR - *Subscription Profile Repository*). No entanto, na maior parte das vezes isso não chega, pois é necessário, por exemplo, perceber o estado de congestão de determinados pontos da rede (e.g. células/eNBs), quem é que está a congestionar a rede, e o que deve ser feito para resolver o problema, e.g. eliminar alguns subscritores de menor valor, os que consomem mais recursos (BitTorrents), ou reduzir um pouco a alguns ou a todos, de forma a ajustar os consumos aos recursos disponíveis. As ações deverão ser tanto mais vigorosas quanto maior for a congestão e a premência por reduzi-la. As situações serão tanto mais eficazes quanto antes sejam tomadas medidas. Idealmente, medidas preventivas são mais eficientes do que medidas reativas, e medidas com baixos níveis de congestão têm menos impactos nos subscritores do que medidas com altos níveis de congestão.

Este trabalho teve como principais objetivos estudar a problemática do controlo de congestão, assim como de dotar o componente PCRF da PT Inovação (ip-Raft) de um mecanismo deste tipo. O artigo começa por descrever o estado da arte no controlo de congestão nos principais concorrentes, passando seguidamente para a identificação dos problemas mais comuns. Posteriormente, são identificados alguns casos de uso, assim como as políticas (algoritmos) que seriam adequadas para a resolução desses problemas genéricos. Seguidamente, são descritas a arquitetura e implementação desenvolvidas, assim como a prova de conceito realizada em ambiente laboratorial. Finalmente, o artigo conclui.

2. ESTADO DA ARTE

Tradicionalmente, as redes 3GPP, e em particular as LTE, têm mecanismos de controlo de admissão (CAC) nas suas redes. Este mecanismo regula a admissão de tráfego de dados nas comunicações, garantindo um determinado nível de QoS por tipo de serviço/subscritor, ou um determinado desempenho dos elementos de rede.

O eNB (célula LTE) é o principal responsável pela implementação de alguns dos processos normalizados de CAC. Usando os limites de tráfego definidos pelo PCRF para cada subscritor/serviço e com base no QCI (QoS Class Indentifier), este componente gere os recursos rádio, fazendo o escalonamento e priorização de tráfego, alocação de recursos e constante monitorização do seu uso. O eNB controla também o congestionamento através do parâmetro ARP (*Allocation and Retention Priority*), o qual faz com que alguns subscritores/serviços sejam eliminados quando episódios de congestão acontecem.

Estes mecanismos têm apenas em consideração o estado do eNB em particular, pelo que não tem em consideração o estado global da rede, a existência de redes alternativas, ou o perfil do subscritor e as necessidades de um serviço em particular. Assim, as decisões tomadas podem não ser as mais interessantes em cada momento para o operador. Por essa razão, existe cada vez a necessidade de fornecer um controlo mais fino e mais inteligente, de forma a corresponder mais fielmente aos interesses dos operadores e dos seus clientes.

Hoje em dia já existem na indústria algumas soluções que podemos dividir em dois tipos: (1) os que integram mecanismos de monitoria da rede, usando sistemas de *probing* ou através do *reporting* do O&M dos equipamentos (eNB, ou outros), de forma a permitirem atuar dinamicamente na largura de banda alocada (que classificamos de RAN-aware); e (2) os que usam estimativas ou histórico de uso non-RAN-aware. A tabela a seguir resume algumas das soluções já existentes do mercado.

FABRICANTE	CARACTERÍSTICAS	DESCRIÇÃO
Sandvine	3 passos: medir, gerir, otimizar Interligação com HSS e SPR	Atua como PCRF, pode usar dados de probes passivas, RAN-aware.
Tekelec	Interligação com HSS e SPR	Atua como PCRF, pode usar dados de probes passivas, RAN-aware.
Openet	Baseadas em sazonalidade e calendário	Non - RAN-aware
Movik	Capacidade de rede heterogénea.	RAN-aware
Bridgewater Systens/Amdocs	Informação dada pelas estações base.	Usos de uma janela temporal para permitir, Non-RAN-aware
Alcatel	Correlação domínio IP com domínio rádio para estimativas.	Uso de uma entidade gestora da rede (monitoria)

Tabela 1. Tabela de solução CAC comerciais

3. PROBLEMAS E SOLUÇÕES

Em situações de congestão, seja ela mais ou menos grave, os subscritores são sujeitos a comportamentos por omissão definidos nas normas 3GPP: QCI e ARP acima descritos. No entanto, estas abordagens são estáticas (definidas no perfil do subscritor) e não têm em conta o contexto específico do subscritor, nomeadamente, o estado da rede, outros subscritores que também estão a sofrer de congestão, o perfil do subscritor (de maior ou menor valor), o tipo de serviço (dados puros, voz, ou vídeo), ou o estado financeiro do cliente (saldo, faturas não pagas, etc.). O ideal seria que todos (ou parte) destes parâmetros pudessem ser tidos em conta na decisão de quem dever ser mais e menos penalizado.

Por outro lado, após ter o conhecimento da informação de contexto, existem várias hipóteses de atuação possíveis. Em geral, estas passam por rejeitar um ou mais subscritores já com sessão estabelecida, não permitir a entrada de novos subscritores, reduzir a largura de banda atribuída (via perfil) a todos ou a um subconjunto dos subscritores acometidos pela congestão, ou o redireccionamento de alguns subscritores para outras tecnologias de rede, se disponíveis, e.g. Wi-Fi. Estas são só algumas das possibilidades, mas existem muitas mais e todas elas merecem ser exploradas.

De forma a evitar alterações bruscas na rede, é importante definir limiares de atuação, de forma a tomar decisões suaves quando o nível de congestão se aproxima (mas ainda está acima) dos limiares considerados como início de congestão, passando a decisões mais enérgicas quando o nível de congestão se aproxima de valores que claramente impõe restrições substanciais aos subscritores e provoca uma diminuição significativa da qualidade de experiência geral, redundando em prejuízo para todos.

Com vista a encontrar uma solução para o problema, pretende-se construir um mecanismo de CAC (*Call Admission Control*), associado ao PCRF, que use um "*real-time feedback loop*" que permita examinar e ajustar continuamente os recursos de rede às necessidades, permitindo ao operador adaptar-se rapidamente a episódios de congestão. Este processo será composto essencialmente por três passos.

- **Analisar** a utilização dos recursos na rede e restante informação de contexto, por forma a detetar os casos de congestão (antes ou depois de eles acontecerem);
- **Decidir** a ação a tomar quando ocorrem episódios de congestão, quer provocada pelo aumento do consumo dos recursos dos subscritores em sessão, quer pela chegada de novos subscritores (estabelecimento inicial ou *handover*);
- **Atuar** na rede fazendo o *enforcement* das decisões, para prevenir episódios de congestionamento, ou minorar as suas consequências quando estes aconteçam.

No estudo da aplicação de mecanismos de CAC, têm que ser considerados 2 cenários: (1) cenário com PCEF (*Policy and Charging Enforcement Function*), colocalizado no P-GW, e (2) cenário com TDF (*Traffic Detection Function*), por vezes denominado DPI (*Deep Packet Inspector*), e colocado à saída do P-GW.

No cenário (1), o PCRF e atua na rede através da interface Gx, que liga o PCRF ao PCEF/P-GW. Neste cenário, o PCRF pode intervir na negociação dos *bearers*, alterando alguns parâmetros – como QCI, ARP, MBR ou GBR – ou simplesmente recusando o estabelecimento de um novo *bearer* (sessão). No cenário (2), o PCRF atua na rede através da interface Sd (semelhante à Gx, mas mais limitada), que o liga o PCRF ao TDF. Neste cenário, o PCRF interage com um elemento (TDF) que não tem qualquer interação com o P-GW, e não participa diretamente no fluxo de criação de uma sessão. Dessa forma, o PCRF não tem a capacidade nem de condicionar a criação de novos *bearers* (sessões), nem de alterar os seus principais parâmetros de acordo com determinadas políticas. Resta-lhe a capacidade de controlar o tráfego à saída do P-GW, limitando-se a condicionar o tráfego.

4. CASOS DE USO

CENÁRIO 1 - CONGESTIONAMENTO DEVIDO A UM NOVO SUBSCRITOR/SERVIÇO

Caso de uso 1-I: Admissão de um subscritor/serviço prioritário

Descrição: Chega um pedido de admissão à rede de um subscritor/serviço prioritário e a célula (eNB) onde o UE está ligado encontra-se totalmente congestionada, com os *bearers* lá existentes a consumirem todos os recursos a que foram autorizados.

Soluções Possíveis: (1) Reduzir largura banda de subscritores/serviços menos prioritários até obter a largura de banda necessária; (2) Reduzir largura de banda de subscritores/serviços não prioritários e prioritários; (3) Eliminar a sessão de alguns subscritores/serviços não prioritários; (4) *Handover* de subscritores/serviços não prioritários para Wi-Fi, se possível.

>>>

Caso de uso 2-II: Admissão de um subscritor/ serviço não prioritário

Descrição: Chega um pedido de admissão à rede de um subscritor/serviço não prioritário e a célula (eNB) onde o UE se encontra está totalmente congestionada com os *bearers* lá existentes a consumirem todos os recursos a que foram autorizados.

Soluções Possíveis: (1) Não aceitar o estabelecimento do *bearer* do novo subscritor; (2) Reduzir largura banda de subscritores/serviços menos prioritários até obter a largura de banda necessária; (3) Eliminar a sessão de alguns subscritores/serviços não prioritários; (4) *Handover* de subscritores/serviços não prioritários para Wi-Fi, se possível.

CENÁRIO 2 - CONGESTIONAMENTO DEVIDO A *OVERPROVISION/HANDOVER*

Caso de uso 2-I: Aumento do consumo de um subscritor

Descrição: Um *bearer* começa a consumir completamente os recursos a que tem direito e congestiona a célula, sendo que os restantes *bearers* lá existentes não estão a consumir tudo aquilo a que foram autorizados, implicando o CAC ter informação atualizada sobre o tráfego de cada *bearer*. O *bearer* que causa o congestionamento pode consistir num subscritor mais ou menos prioritário.

Soluções Possíveis: (1) Reduzir largura banda de subscritores/serviços menos prioritários até obter a largura de banda necessária; (2) Reduzir largura de banda de subscritores/serviços não prioritários e prioritários; (3) Eliminar a sessão de alguns subscritores/serviços não prioritários; (4) *Handover* de subscritores/serviços não prioritários para Wi-Fi, se possível.

Caso de uso 2-II: *Handover* de um subscritor/serviço prioritário

Descrição: Ocorre um *handover* que congestiona a célula, sendo que os *bearers* lá existentes não estão a consumir tudo aquilo a que foram autorizados, implicando o CAC ter informação atualizada sobre o tráfego de cada *bearer*. O subscritor que causa o congestionamento é um subscritor/serviço prioritário.

Soluções Possíveis: (1) Reduzir largura banda de subscritores/serviços menos prioritários até obter a largura de banda necessária; (2) Reduzir largura de banda de subscritores/serviços não prioritários e prioritários; (3) Eliminar a sessão de alguns subscritores/serviços não prioritários; (4) *Handover* de subscritores/serviços não prioritários para Wi-Fi, se possível.

Caso de uso 2-III: *Handover* de um subscritor/ serviço não prioritário

Descrição: Ocorre um *handover* que congestiona a célula, sendo que os *bearers* lá existentes não estão a consumir tudo aquilo a que foram autorizados, implicando o CAC ter informação atualizada sobre o tráfego de cada *bearer*. O subscritor que causa o congestionamento é um subscritor/serviço não prioritário.

Soluções Possíveis: (1) Não aceitar o *handover* do novo subscritor; (2) Reduzir largura banda de subscritores/serviços menos prioritários até obter a largura de banda necessária; (3) Eliminar a sessão de alguns subscritores/serviços não prioritários; (4) *Handover* de subscritores/serviços não prioritários para Wi-Fi, se possível.

CENÁRIO 3 - DESCONGESTIONAMENTO DE UM ENB

Caso de uso 3-I: *Handovers* para outros eNBs

Descrição: Este caso de uso acontece na sequência de um *handover* que congestionou a célula, tendo sido feito um *downgrade* dos subscritores/serviços não prioritários e/ou prioritários. Neste caso de uso alguns subscritores fizeram *handover* para outros eNBs e o eNB deixou de estar congestionado, sendo assim necessário efetuar o *upgrade* dos serviços anteriormente *downgraded*, para repor a situação inicial de pré-congestão. Os subscritores que descongestionaram a célula podem ser subscritores prioritários ou não prioritários.

Soluções Possíveis: (1) Aumentar largura banda de subscritores/serviços mais prioritários, mantendo a rede liberta; (2) Aumentar largura banda de subscritores/serviços menos prioritários; (3) *Handover* de subscritores/serviços não prioritários de Wi-Fi para rede móvel, se possível.

5. POLÍTICAS CAC

DIAGRAMAS DE FLUXO

O algoritmo pode ser invocado de duas formas diferentes, quer quando um novo pedido chega ao PCRF, quer quando é detetado (via OSSs) que a rede está congestionada. O resultado destas políticas pode afetar apenas o causador do evento ou os restantes subscritores da célula, A Figura 2 mostra o diagrama de fluxo da política.

>>>

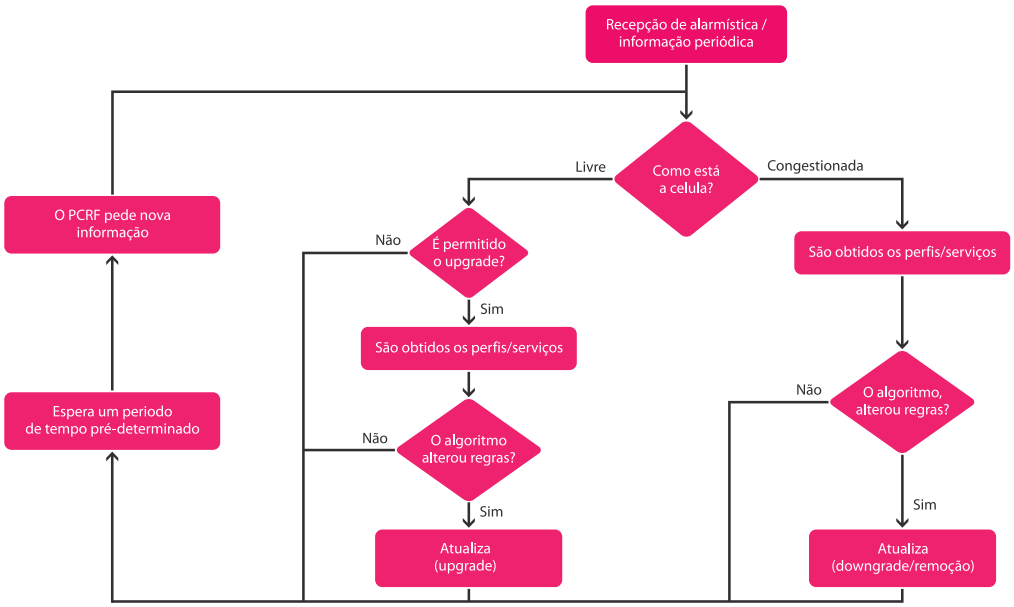


Figura 2. Fluxograma do "Real Time Feedback Loop"

INPUTS/OUTPUTS CAC

Os principais parâmetros de entrada (*inputs*) que o mecanismo CAC utiliza para decidir as políticas a aplicar e as ações de saída (*output*) são apresentadas de seguida.

Inputs:

- Nível de congestão por célula - percentagem da largura de banda da célula ocupada;
- Consumo por subscritor/serviço - consumo de cada *bearer* de cada subscritor;
- Estatística de congestão - informação histórica sobre a congestão da célula;
- Tipo de subscritor/serviço - se é um subscritor *premium*, gold serviço Internet, VoIP;
- Atributos de cada *bearer* - largura de banda atribuída, reservada;
- QCI atual - classe de QoS atribuída;
- Redes disponíveis para *handover* - outras redes na proximidade (e.g. Wi-Fi, 3G).

Outputs:

- *Downgrade*/*upgrade*/bloqueio - Alteração dos recursos dos subscritores/serviços;
- Reduzir/remover o serviço X - atuar no tipo de serviços X (e.g. P2P);
- Limitar os serviços no tempo - atuar enquanto a célula estiver congestionada;
- Reduzir/remover serviços mais consumidores - serviços que congestionam a célula;
- Dar prioridade a serviços de tempo real - VoIP, Vídeo Conferência, etc.;
- Limitar os subscritores "abusadores" - que estejam a usar intensivamente a célula;
- Priorizar subscritores/serviços *premium* - subscritores/serviços mais importantes;
- Alterar QCIs em casos de congestão - alterar para QCIs com menos (mais) recursos;
- *Handover* para outras redes - mover o subscritor para outras redes (Wi-Fi).

ALGORITMOS E EXEMPLOS

O algoritmo CAC pode ser caraterizado por dois tipos de atuações: (1) o *downgrade* (restrição dos parâmetros de QoS dos serviços contratualizados) e o *upgrade* (retoma dos serviços aos parâmetros de QoS contratualizados). De seguida são apresentados alguns exemplos que contemplam os dois tipos de atuações.

Algoritmo:
As regras a serem aplicadas variam com o nível de congestionamento da célula. Para um caso de 3 níveis de congestionamento (Td1, Td2 e Td3) para *downgrade* (em que Td1<Td2<Td3), e para 1 nível de congestionamento (Tu1) de *upgrade*, podem ter-se as seguintes regras e os seguintes valores

$$Td1=70\% < Td2=80\% < Td3=90\% \,;\, Tu=50\%$$

teremos:

a) $70\% \leq$ Ocupação $< 80\% \rightarrow$ *Downgrade* de serviços/subscritores menos prioritários;
b) $80\% \leq$ Ocupação $< 90\% \rightarrow$ *Downgrade* de todos os serviços/subscritores;
c) $90\% \leq$ Ocupação $\rightarrow$ Eliminação de serviços/subscritores.
d) Ocupação $< 50\% \rightarrow$ *Upgrade* de todos os subscritores/serviços.

>>>

Estes valores podem ainda ser afinados. Por exemplo, se o histórico de congestão mostrar que a célula costuma estar congestionada neste dia ou hora, os limiares devem ser diminuídos, de acordo com um delta fornecido pelo operador.

No caso do *downgrade*, a ordem de atuação para cada intervalo é a seguinte:

1. *Downgrade* aos subscritores/serviços que usam recursos acima do contratado;
2. *Downgrade* aos subscritores/serviços "abusadores" (intensivos);
3. *Downgrade* nos serviços de baixo valor (e.g. P2P);
4. *Downgrade* nos serviços que consumam muitos recursos;
5. *Downgrade* nos restantes serviços/subscritores.

No caso do *upgrade*, a ordem de atuação para cada intervalo (neste caso 1) é a seguinte:

1. *Upgrade* dos subscritores/serviços mais prioritários;
2. *Upgrade* dos subscritores/serviços menos prioritários;
3. *Upgrade* dos subscritores/serviços para valores acima do contratado.

6. ARQUITETURA E IMPLEMENTAÇÃO

ARQUITETURA

O mecanismo CAC é uma funcionalidade que faz parte do PCRF como mostra a Figura 3.

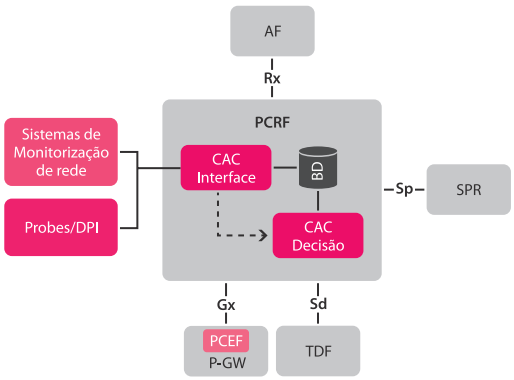


Figura 3. Arquitetura da integração do CAC com o PCRF

Arquitetura Geral

- **Base de Dados** - É interna ao mecanismo CAC, e contém toda a informação de parametrização e configuração necessária ao mecanismo CAC;
- **CAC Interface** - É responsável por recolher e processar toda a informação obtida a partir de sistemas externos (e.g. sistemas de monitorização de rede, sistemas de *Probing*, ou DPIs, entre outros), que alimentam os mecanismos CAC. Armazena a informação na Base de Dados e gera *triggers* para o CAC Decisão se necessário;
- **CAC Decisão** - É responsável por tomar todas as decisões, consoante a informação que recebe (*inputs*), atualizar a base de dados com as alterações efetuadas e gerar decisões (*outputs*). O seu motor de decisão é o algoritmo descrito anteriormente.

IMPLEMENTAÇÃO

O mecanismo CAC foi implementado no PCRF da PT Inovação, produto ipRaft, que é uma solução de Policy para redes de dados. A tecnologia usada para este desenvolvimento está alinhada com as restantes capacidades deste produto, sendo descrita abaixo por módulo.

CAC Decisão

O módulo CAC Decisão foi implementado em QRE (*Quantum Rules Engine*) e Java. O QRE é um motor de avaliação de regras da PT Inovação, que permite configurar e avaliar regras (políticas), tornando possível que estas possam ser configuradas fora dos processos de negócio sem que seja necessário recompilar o código sempre que é alterado, aumentando a sua flexibilidade. Em Java encontram-se todas as funções mais específicas e complexas, como o *Downgrade*, *Upgrade*, *Drop* e toda a interação com a base de dados.

Base de Dados

A base de dados foi implementada à semelhança do SPR, em Java, de acordo com a sua especificação (já descrita anteriormente). Esta foi implementada de forma a otimizar as pesquisas durante a realização das funções complexas em Java, e despoletadas pelo QRE.

CAC Interface

Este módulo não chegou a ser implementado, apenas foi especificado, uma vez que não houve tempo para isso. Este terá de ser trabalhado no futuro, quando for possível obter informação vinda de sistemas externos (e.g. monitoria, *probing*, alarmística, etc.).

7. PROVA DE CONCEITO

O cenário geral usado para efetuar a prova de conceito é apresentado na Figura 4.

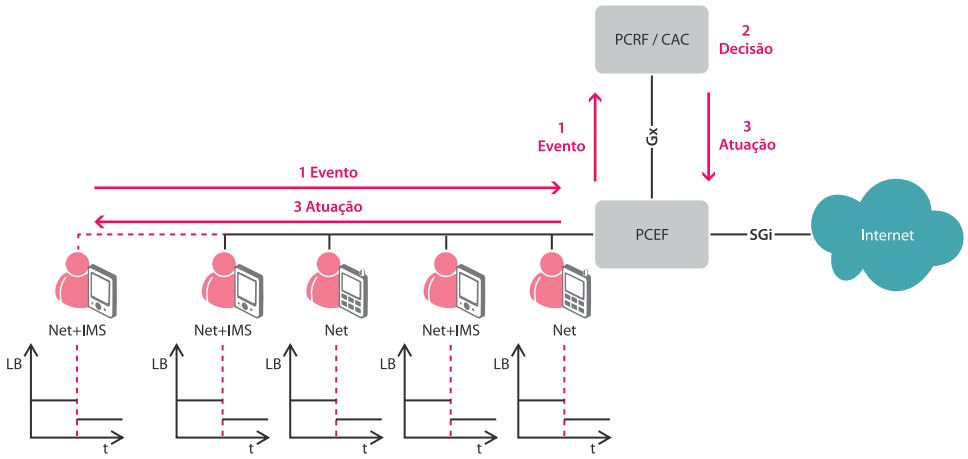


Figura 4. Especificação do demonstrador, com um cenário genérico

Este consiste num conjunto de subscritores com diferentes perfis, ligados através de um PCEF (Cisco SCE 8000), por uma interface com largura de banda limitada. Por omissão, estes subscritores consomem o máximo de *throughput* que conseguirem congestionando a interface. A entrada na rede de um novo subscritor (simulando a criação de um *bearer*) despoleta um evento para o PCRF (via interface Gx) e este, através do mecanismo CAC, decide em que condições o cliente se pode ligar tendo em conta: o tipo de perfil, o estado atual do acesso e o tipo de perfil dos subscritores que nele se encontram.

Cada perfil de subscritor possui um mapeamento de características no PCEF, que permite ao PCRF condicionar o acesso dos subscritores de acordo com um algoritmo configurável. Este algoritmo de decisão tem como *inputs*: o estado de congestão da célula e os *thresholds* configurados para esta. A tabela seguinte (Tabela 2) apresenta o mapeamento das condições/ações que o algoritmo tem em conta.

Nível de Congestão (n)	Ação
Down T1 < n < Down T2	Downgrade de 1 nível;
Down T2 < n < Down T3	Downgrade de 2 níveis;
n > Down T3	Bloqueio
n < Up T0	Upgrade

Tabela 2. Tabela de limiares de congestão

A afetação de cada sessão é feita de forma faseada, isto é, nem todos os subscritores com o mesmo perfil são considerados, sendo que após o *threshold* ser atingido novamente, o algoritmo deixa de atuar. Caso a célula se encontre congestionada, são primeiro analisados os perfis menos prioritários e quando a célula volta a ficar descongestionada são analisados primeiro os perfis mais prioritários, com vista a repor o seu perfil inicial. Para um melhor entendimento do mecanismo vejam-se os exemplos a seguir.

DEMO 1 - NOVO SUBSCRITOR EM CÉLULA CONGESTIONADA I

A entrada de um novo subscritor (S3) congestiona a célula, sendo que os subscritores que lá se encontram (S1 e S2) estão a consumir tudo a que têm direito (com base no seu perfil). O subscritor (S3), que causa o congestionamento é aceite; no entanto, fica com o seu perfil *downgraded* juntamente com o subscritor menos prioritário (S1). Posteriormente, quando sai um subscritor (S2), a célula deixa de estar congestionada e é feito o *upgrade* dos subscritores para o débito contratado associado ao respetivo perfil.

DEMO 2 - NOVO SUBSCRITOR EM CÉLULA CONGESTIONADA II

A entrada de um novo subscritor (S3) congestiona a célula, sendo que os subscritores que lá se encontram (S1 e S2) estão a consumir tudo a que têm direito (com base no seu perfil). O subscritor (S3) que causa o congestionamento é aceite condicionalmente, em detrimento dos subscritores de perfis menos prioritários que são bloqueados (S2), exceto os que têm os serviços especiais (S1) e não podem sofrer condicionamentos. Posteriormente, quando sair um subscritor e a célula deixa de estar congestionada, é feito o *upgrade* do subscritor que tinha sido bloqueado (S2) para o tipo de perfil contratado.

DEMO 3 - VARIAÇÃO DO CONSUMO DE CADA SUBSCRITOR

Um subscritor (S1) aumenta o seu consumo e congestiona a célula, sendo que os subscritores que se encontram na mesma célula não estão a consumir tudo a que têm direito. Quando o CAC é notificado, os subscritores menos prioritários (S2 e S3) são *downgraded*. Posteriormente, quando o subscritor (S1) diminui o seu consumo e a célula deixa de estar congestionada, é feito o *upgrade* dos subscritores que foram *downgradeded*.

>>>

RESUMO

	Subscritores na célula	Novo Subscritor	Largura Banda por Perfil	Limiares	Capac. da célula	Decisão I	Decisão II
Demo 1	S1 (Silver) S2 (Gold)	S3 (Gold)	Gold = 4 Mb Silver = 2 Mb Bronze = 1 Mb	Down T0 = 65% Down T1 = 70% Down T2 = 80% Down T3 = 90% Up T0 = 50% Up T1 = 30%	13 Mb	S1 (Silver->Bronze) S2 (Gold) S3 (Silver)	S2 Out S1 (Bronze->Silver) S3 (Silver->Gold)
Demo 2	S1 (Bronze*) S2 (Silver)	S3 (Gold)			7 Mb	S1 (Bronze*) S2 (Silver->Blocked) S3 (Silver)	S3 Out S2 (Blocked->Silver) S3 (Silver)
Demo 3	S1 (Gold) S2 (Silver) S3 (Silver)	S1 (Gold)			10 Mb	S1 (Gold) S2 (Silver->Bronze) S3 (Silver->Bronze)	S1 (Gold) S2 (Bronze->Silver) S3 (Bronze->Silver)

Tabela 3. Tabela Resumo de Demonstrações

8. CONCLUSÕES

Este artigo descreve os resultados do projeto PI 2012/2013 "PCRF", o qual pretendia estudar, especificar e implementar mecanismos de controlo (CAC) que permitissem gerir o congestionamento em redes móveis LTE. Para isso, foram definidos e estudados um conjunto de casos de uso, de forma a demonstrar que, com a informação de perfil dos subscritores e o auxílio de indicadores vindos da rede, é possível criar mecanismos que permitam dar uma melhor experiência de utilização, e ao mesmo tempo reduzir custos no operador, através do uso mais eficiente dos recursos.

Estes mecanismos permitem tomar as melhores decisões em períodos de congestão, fazendo com que os clientes de maior valor ou os serviços mais sensíveis (voz, vídeo) sejam protegidos face aos restantes, em situações em que os recursos disponíveis não são suficientes. Por outro lado, com uma gestão adequada dos recursos, é possível tomar ações preventivas de forma a evitar que a congestão chegue a acontecer, atuando logo que existam alguns indícios ligeiros de congestão.

O projeto foi desenvolvido em várias fases, começando por ser efetuado um estudo sobre o estado da arte no controlo de admissão e reserva de recursos. Seguidamente foram identificadas as métricas e indicadores necessários ao mecanismo, olhando também para algumas soluções que alguns dos mais fortes competidores nesta área apresentam. Foi nesta fase que se reforçou a perceção da importância dos sistemas externos na obtenção de indicadores do estado da rede em "tempo real".

Na fase posterior, começou por identificar-se as políticas passíveis de serem implementadas no mecanismo CAC a desenvolver, ao mesmo tempo que se especificavam e afinavam os algoritmos mais adequados. Foram assim obtidas um conjunto de possíveis políticas, os *inputs* necessários para tomar decisões, assim como definidos os *outputs* que seriam acionados na rede.

Na fase de implementação, foi usada a plataforma tecnológica do ipRaft, tendo sido aí especificado e implementado um mecanismo de CAC, integrando-o com a plataforma existente, de forma a que facilmente este trabalho pudesse ser adicionado ao produto.

Na última fase, foi especificado o demonstrador do mecanismo, definindo um conjunto de cenários representativos para a prova de conceito (PoC). Foi então efetuada a integração do mecanismo desenvolvido com os componentes de rede reais (P-GW/PCEF Cisco). Para o PoC optou-se por utilizar um programa de monitorização de largura de banda em vez de UEs (terminais) reais, de forma a melhor poder avaliar os resultados obtidos pelo mecanismo CAC.

Em resumo, este projeto possibilitou a inclusão de funcionalidade no ip-Raft (PCRF) que permite a diferenciação do serviço em situações de congestionamento, assim como uma melhoria da qualidade global percebida pelos subscritores (QoE). Por outro lado, esta capacidade materializa-se numa redução dos custos nos operadores, através de uma utilização mais eficiente dos recursos.

Como trabalho futuro, fica por implementar um módulo que permita a receção e agregação de toda a informação recebida de sistemas externos. Quanto mais completa seja essa informação, mais eficaz será o mecanismo, e mais eficiente será a gestão dos recursos. Em estudo ainda, está o mecanismo que permite despoletar o *handover* de terminais para outras tecnologias (e.g. Wi-Fi), integrando o PCRF com outros componentes, nomeadamente o ANDSF (Access Network Discovery and Selection Function). Estes trabalhos estão já a ser endereçados no PI 2013/2014 dedicado também ao PCRF.

CVS DOS AUTORES

Carlos Marques, concluiu o seu Mestrado Integrado em engenharia Eletrónica e de Telecomunicações no Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro em Julho de 2010. Foi bolseiro de investigação científica no Instituto de Telecomunicações, localizado no campus da Universidade de Aveiro, de 2010 a 2013. Esteve envolvido em alguns projetos nacionais, tais como, o Panorama e outros em colaboração com a PT Inovação, e em alguns projetos europeu, tais como o Euro-NF. Em 2013 iniciou a sua atividade profissional na PT Inovação, localizada em Aveiro, e a partir daí tem estado a trabalhar como investigador/desenvolvedor nas áreas de 3GPP e cloud networking. Tem estado envolvido em alguns projetos europeus, tais como, o MCM. Os seus interesses situam-se nas áreas relacionadas com plataformas de simulação, redes em malha sem fios, contexto, cloud networking e 3GPP.

Carlos Parada obteve a Licenciatura em Engenharia de Sistemas e Informática no ano de 2000, pela Universidade do Minho, em Braga. Em 2009 obteve o grau de Master of Science (MSc) in Information Networking, pela Carnegie Mellon University e pela Universidade de Aveiro (dual-degree). Desde 2000, ano em que começou a trabalhar na Portugal Telecom Inovação, Aveiro, esteve envolvido em projetos de I&D nacionais e europeus nas áreas das redes IP, IPv6, redes de core (MPLS), redes fixas (ADSL/GPON) e redes móveis (3GPP). Também participou em projetos de auditoria e consultoria nas mesmas áreas em operadores como a TMN, a UNITEL, ou a CVT. Ao nível dos produtos, participou em projetos de desenvolvimento de serviços e plataformas para redes fixas e móveis, como UNIBOX, GW-WAP, RADIUS, ou plataformas de entrega de conteúdos, entre outros. Desde 2010, tem assumido funções de investigador e consultor interno nas mais variadas áreas de atuação da empresa.

Carlos Rodrigues, licenciado em Engenharia Electrónica e Telecomunicações pela Universidade de Aveiro, em 2001. Iniciou a sua atividade profissional na Portugal Telecom Inovação em 2001, desempenhando atividades na área das redes IP. Entre 2004 e 2007 participou no desenvolvimento de aplicações e arquiteturas para controlo de sessões de dados em tempo-real. De 2008 a 2012 participou no desenvolvimento de aplicações de Controlo de Acesso à Rede". Desde 2012 é o responsável técnico de desenvolvimento dos produtos DSCP e ipRaft PCRF.

Filipe Rodrigues, concluiu o Mestrado Integrado em Engenharia Eletrónica e Telecomunicações pela universidade de Aveiro. Após o Mestrado, tornou-se bolseiro de investigação do Instituto de Telecomunicações, onde trabalhou em projetos co-financiados pela Fundação para a Ciência e Tecnologia (FCT). Ingressou na PT Inovação no início de 2012 para operacionalizar a plataforma OPEN EPC. Desde então tem participado em desenvolvimentos dos projetos da PTIN para LTE ANDSFnet, CAC LTE e ProbeNI_LTE. Atualmente os seus interesses situam-se na área das comunicações sem fios, rede de acesso e core LTE.

Francisco Fontes, licenciado em Engenharia Electrotécnica e de Computadores, ramo de Electrónica e Telecomunicações, pelo Instituto Superior Técnico (Setembro de 1991) e Doutorado pela Universidade Politécnica de Madrid (Novembro de 2000) na área de gestão distribuída de redes de telecomunicações. Em Setembro de 1991 iniciou a sua atividade profissional na PT Inovação (CET) tendo-se especializado em tecnologias de rede de banda larga. Atualmente os seus interesses situam-se na área das arquiteturas de redes, em especial na sua evolução para arquiteturas RPG All-IP, com ênfase no IMS. É especialista em tecnologias de rede local e acesso, IPv6 e Multicast. Também colabora com a Universidade de Aveiro/DETI, na qualidade de Professor Auxiliar Convidado, para as áreas de redes IP aplicadas às telecomunicações.

Miguel Santos, licenciado em Engenharia Electrotécnica e Computadores pela Faculdade de Engenharia da Universidade do Porto (especialização em Telecomunicações e Computadores) em 1998, concluiu o mestrado em Engenharia Electrotécnica e Computadores na Universidade da Florida em 2001. Desde 2002 trabalha na PT Inovação integrado na equipa de desenvolvimento da solução de redes inteligentes. Desempenha desde 2008 as funções de gestor da divisão no departamento Desenvolvimento de Plataformas de Rede e Soluções Multimédia, e é responsável pelo desenvolvimento de soluções na área do controlo em redes de dados e IMS onde se inclui os sistemas ipRaft (solução de Policy Management), WAP gateway (adaptação de pedidos Wap dentro da rede do operador), ipTiller (solução para controlo de autenticação e autorização de acesso a redes de dados; incluindo as funções de DHCP, AAA e CLF/LRF).

Pedro Neves, Doutorado e Mestre em Engenharia Electrónica, Telecomunicações e Informática pela Universidade de Aveiro em 2012 e 2006, respetivamente. Em 2003 tornou-se bolseiro de investigação do Instituto de Telecomunicações, onde trabalhou em projetos co-financiados pela Comissão Europeia na área de mobilidade e qualidade de serviço. Em Junho de 2006 iniciou atividade na PT Inovação na mesma área de trabalho. Desde 2010 que acumula também atividades de investigação na área de Cloud Computing. Participou em mais de 10 projetos de colaboração internacional, é co-autor de 6 livros internacionais e de mais de 30 artigos publicados em revistas e conferências internacionais.

Susana Sargento, (http://www.av.it.pt/ssargento) obteve o Doutoramento em 2003 em Engenharia Electrotécnica na Universidade de Aveiro. Ela foi docente do Departamento de Ciências de Computadores da Universidade do Porto de Setembro de 2002 a Fevereiro de 2004, e encontra-se na Universidade de Aveiro e no Instituto de Telecomunicações desde Fevereiro de 2004, onde actualmente lidera o grupo de Arquitecturas e Protocolos de Redes (http://nap.av.it.pt). Ela faz parte também do corpo docente convidado do Departamento de Eng. Electrotécnica e de Computadores da Universidade de Carnegie Mellon, USA, desde Agosto de 2008, onde realizou 'faculty exchange' em 2010/2011. Desde Março de 2012, a Susana é co-fundadora de uma empresa de redes veiculares, a Veniam'Works, que tem como objectivo construir uma rede Internet com base em veículos. A Susana tem estado envolvida em vários projectos nacionais e internacionais, tendo liderado algumas actividades, como as actividades de qualidade de serviço e de redes ad-hoc do projecto europeu FP6 IST-Daidalos. Ela esteve recentemente envolvida em vários projectos FP7 (4WARD, Euro-NF, C-Cast, WIP, Daidalos, C-Mobile), projectos nacionais, e projectos do programa CMU|Portugal (DRIVE-IN com a Universidade de Carnegie Mellon). Os seus interesses de investigação centram-se nas áreas de redes de nova geração e redes da Internet do futuro, mais especificamente em QoS, mobilidade, redes auto-geridas e cognitivas. A Susana faz regularmente parte do Painel de Especialistas nos programas de investigação europeus.

Telma Mota, concluiu a Licenciatura e Mestrado em Engenharia Electrotécnica e de Computadores na Universidade do Porto. Ingressou na empresa TLP SA, onde realizou trabalho de planeamento e dimensionamento de redes de comutação digital, Redes Inteligentes e teletráfego. Desde 1994 que integra a PT Inovação e tem estado ligada às áreas de gestão e arquitecturas de Redes e Serviços; IN, evolução da IN, TINA, Parlay, IMS, TISPAN e MBMS, assim como às normas 3GPP que se dedicam a definir aspectos de estabelecimento de Sessões Multimédia, QoS, Mobilidade e Multicast. Recentemente tem-se dedicado às arquitecturas de serviços; OMA, SOA, Web 2.0. Participou em diversos projectos Europeus (Eurescom e IST), liderou o C-CAST e na PTIN é responsável pela divisão "Plataformas e Redes Multiserviço".

Tiago Cardoso, concluiu o Mestrado Integrado em Engenharia Electrotécnica e de Computadores, ramo de Telecomunicações e especialização em Redes e Serviços de Comunicações, na Faculdade de Engenharia da Universidade do Porto. Ingressou na PT Inovação no inicio de 2010 para realizar a sua dissertação intitulada "Mobilidade Entre Diferentes Redes de Acesso em Terminais de Próxima Geração" cujo trabalho se inseriu no âmbito do projecto europeu HURRICANE. Posteriormente esteve envolvido no desenvolvimento da solução MyConnect, gestor de conectividade para PC's, actualmente encontra-se a trabalhar na solução ipRaft (PCRF/CAC/ANDSF) da PT Inovação.

15 Integração Wireless LAN no Core EPC



FILIPE RODRIGUES FRANCISCO FONTES NUNO PALHARES

PALAVRAS CHAVE
3GPP, WLAN, LTE, EPC

A utilização das redes WLAN (Wireless LAN) em conjunto com as redes 3G e 4G/LTE (*Long Term Evolution*[1]), para fazer face às limitações de espectro e aumento do tráfego dos utilizadores, desperta grande interesse por parte dos operadores de telecomunicações. Existem já soluções que implementam mecanismos de *offloading* de tráfego para a rede WLAN, utilizando clientes nos terminais, como por exemplo o MyConnect da PT Inovação (PTIN). Tal consegue-se mudando os terminais de rede, mas sem integração com o núcleo *Evolved Packet Core* (EPC) e de forma não transparente para o utilizador, ou seja, com perda de sessões ativas. Por essa razão, essas soluções são limitativas a uma boa experiência dos utilizadores.

O 3GPP (*3rd Generation Partnership Project*[2]) criou soluções que permitem ao equipamento terminal, de forma transparente, selecionar e mudar de rede de acesso, comutando entre WLAN e redes 3GPP, dessa forma ligando-se à rede de acesso mais vantajosa.

Neste artigo são identificadas e analisadas as soluções presentes nas normas 3GPP e destacadas as soluções bem como os vetores de evolução. São também referidos os impactos e os papéis que alguns componentes de desenvolvimento da PTIN, como o ANDSF (*Access Network Discovery and Selection Function*), PCRF (*Policy and Charging Rules Function*) e OLT-WZG (*Optical Line Termination – Wireless Zone Gateway)*, podem ter.

[1] http://www.3gpp.org/lte
[2] http://www.3gpp.org

1. INTRODUÇÃO

A rede *Long Term Evolution* (LTE), vulgarmente referida como quarta geração móvel ou 4G, é uma realidade e passa por ela o futuro de curto e médio prazo das telecomunicações sem fios, com a evolução para o *LTE Advanced*.

Por outro lado, a tecnologia WLAN tem um evidente sucesso que vai além da utilização no domínio das redes privadas, de âmbito doméstico ou empresarial. Numa época em que a maioria dos dispositivos móveis vem equipada com tecnologia WLAN, várias instituições, como cafés e hotéis, adotaram e disponibilizam *hotspots* para acesso à Internet. Esta adoção foi ainda potenciada pelo facto das redes dos operadores móveis, adotando tecnologia GPRS, disponibilizarem baixos débitos, com coberturas geográficas limitadas e preços elevados. A evolução para a tecnologia 3G, como o *High-Speed Downlink Packet Access* (HSDPA), *High-Speed Packet Access* (HSPA) e, HSPA+, e para o 4G vieram inverter a situação, parecendo, numa primeira fase, que o âmbito de utilização das redes WLAN voltaria a ficar limitado ao domínio privado. A crescente generalização dos chamados *smarthphones* alterou em muito os hábitos de consumo de conteúdos, levando a um crescimento de tráfego por utilizador, evidenciando a necessidade de uma nova fase na coexistência das redes 3GPP e WLAN, com estas a integrarem-se e a complementarem-se de forma a proporcionar a melhor conectividade ao menor custo.

Para o operador, as redes WLAN são vistas como uma oportunidade, principalmente pela sua instalação relativamente fácil e ágil e com um *total cost of ownership (TCO)* inferior às redes 3G e 4G. Esse é o objetivo imediato das soluções de *WLAN offloading*, com a libertação de recursos das redes 3G/4G para serviços mais exigentes, em especial quando a qualidade de serviço (QoS) é uma necessidade. Para tirar total partido desta complementaridade, interessa que o terminal possa ser utilizado para o acesso simultâneo a diferentes serviços, podendo mover-se geograficamente, mudando dinamicamente, de forma transparente e contínua, as ligações necessárias, sobre as diferentes redes de acesso (WLAN e 4G), isolada ou conjuntamente. Para isso é necessária interoperabilidade e mobilidade. É neste contexto que operadores, fabricantes e grupos de normalização estão a propor soluções que integram no núcleo EPC as redes WLAN. Na seleção da melhor solução, a simplicidade e o custo são fatores importantes. Interessa adotar protocolos suportados pela esmagadora maioria dos terminais e sistemas operativos, e que sejam já do domínio do operador, aproveitando assim a experiência e equipamentos já existentes.

2. A ARQUITETURA EPC

O *System Architecture Evolution* (SAE) é a evolução da arquitetura da rede de núcleo 3GPP que, dando suporte ao *Long Term Evolution* (LTE) e com o *Evolved Packet Core* (EPC), originam o *Evolved Packet System* (EPS) [1], conforme a Fig. 1. Nesta evolução, o protocolo IP foi o eleito para o transporte de todos os serviços, deixando-se cair a componente de circuitos (CS – *Circuit Switching*). Como consequência, serviços tradicionais do domínio CS, como a voz e SMS (*Short Message System*), são prestados sobre pacotes, usando o IMS (*IP Multimedia Subsystem*), em particular o chamado *Voice over LTE* ou VoLTE.

Os elementos principais da arquitetura EPC, são evoluções dos elementos da anterior geração, identificando-se: *evolved Node B* (eNB), *Serving Gateway* (SGW), *PDN Gateway* (PGW), *Mobility Ma-*

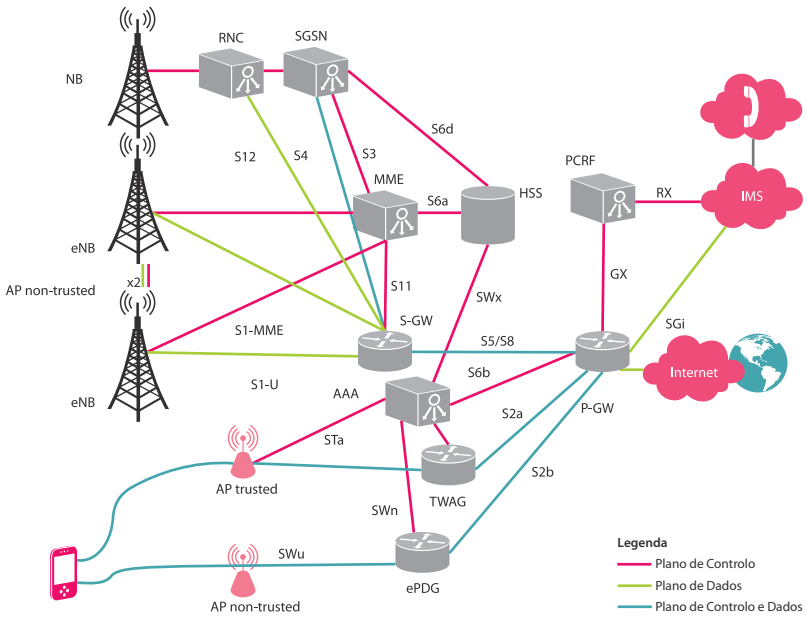


Figura 1. Arquitectura EPC, extendida a redes não 3GPP

nagement Entity (MME) e *Home Subscriber Server* (HSS). Para além destes elementos, devem-se considerar ainda o ANDSF e PCRF como elementos de controlo da rede.

Adicionalmente ao suporte de acessos LTE, rede de pacotes da *GSM EDGE Radio Access Network* (GERAN)) e *Universal Terrestrial Radio Access Network* (UTRAN), a arquitetura evoluiu para integrar outros tipos de acessos, ditos não 3GPP, sendo adicionados os elementos *enhanced Packet Data Gateway* (ePDG) e *Trusted Wireless Access Gateway* [2].

Um diagrama de blocos e forma de interligação de todos esses elementos, está também representado na Fig. 1.

A integração de redes WLAN num núcleo EPC deve responder às seguintes necessidades, de forma a melhorar a experência do utilizador:

- Descoberta automática e seleção. A rede deve possuir mecanismos para assegurar que o equipamento do utilizador (UE) consegue automaticamente descobrir e selecionar entre redes 3GPP e redes WLAN, idealmente de forma transparente para o utilizador.
- Eliminação da necessidade de logins manuais, com a intervenção do utilizador.
- Ligação *Over-The-Air* segura, como em redes 3GPP.
- Mobilidade *seamless* e sem a intervenção do utilizador: assegurar a não interrupção das sessões e sem intervenção do utilizador.
- Controlo de tráfego, com políticas de gestão do mesmo, por forma a evitar congestões.
- *Roaming* transparente: simplificar a experiência do utilizador quando em *roaming*, providenciando conectividade WLAN, com todas as funcionalidades habituais das redes 3GPP.
- Integração com sistemas de tarifação.
- Conectividade ubíqua: assegurar que o utilizador obtém as melhores ligações para as suas necessidades, usando a rede mais vantajosa, sem uma intervenção manual e consciente.

2.1 REDES WLAN

A integração de redes não 3GPP no EPC está definida na TS 23.402 [2] e distingue dois tipos de acessos: os confiáveis (TWLAN - Trusted WLAN) e os não confiáveis. Ser ou não confiável não é uma característica da tecnologia mas antes dependente da relação existente entre o operador do núcleo EPC e o operador da outra rede, bem como dos níveis de segurança proporcionados. Uma rede WLAN confiável normalmente pertence ao operador, com encriptação na interface rádio e um método de autenticação seguro, características estas presentes nas redes 3G e 4G definidas pelo 3GPP. A maioria das atuais soluções de integração de redes WLAN e 3GPP aplicam-se a redes WLAN confiáveis.

2.2 ASPETOS FUNDAMENTAIS DA INTERLIGAÇÃO WLAN-EPC

Neste ponto vão ser descritas as soluções encontradas para os requisitos de integração descritos anteriormente.

2.2.a Continuidade de sessão entre redes 3GPP e WLAN

Existem algumas soluções que providenciam mobilidade entre diferentes redes sem a quebra de sessão IP, ou seja, *seamless* para o utilizador. Tipicamente são classificadas em:

- *Host Based Mobility* (HBM)

Requerem um agente de mobilidade na rede e suporte especial no terminal, para além do interface S2c [4]. Por essa razão não são aqui analisados. Exemplos de protocolos aplicáveis são o *Mobile IP (MIP)* [3] e o *Dual Stack Mobile IP (DSMIP)*.

- *Network Based Mobility* (NBM)

A mobilidade usando funções de rede é preferencial, dado que requer apenas funcionalidades acrescidas na rede, tornando os terminais agnósticos aos mecanismos de mobilidade adotados.

Exemplos de protocolos utilizados neste âmbito são os *Proxy Mobile IP* (PMIP) [5] e *GPRS Tunneling Protocol* (GTP) [6].

Cenário S2a - *Trusted WLAN* (TWLAN)

Para a mobilidade entre redes 3GPP e acessos TWLAN é considerada sempre a existência de uma *gateway* adicional, embora não necessitando do ePDG. A autenticação do utilizador é feita pelo AAA usando as credenciais do *Universal Subscriber Identity Module* (USIM) que são entregues via o interface SWx pelo HSS, sendo baseada em *Extensible Authentication Protocol-subscriber-AKA* (EAP-AKA). Depois é criado um túnel *GPRS Tunnelling Protocol* (GTP) entre o *Trusted Wireless Access Gateway* (TWAG) e o PGW como mostra a figura 2. Os dados do utilizador são transportados entre o AP e o TWAG recorrendo a um túnel GRE (*Generic Routing Encapsulation*).

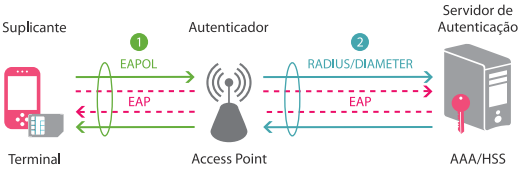


Figura 2. Plano de controlo para mobilidade S2a

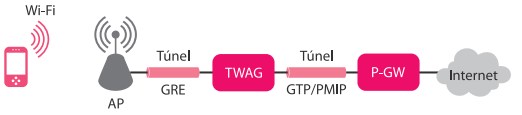


Figura 3. Cenário de acesso TWLAN (plano de dados para mobilidade S2a)

Cenário S2b non-Trusted WLAN

Para redes não confiáveis é utilizado o *enhanced Packet Data Gateway* (ePDG), onde é criado um túnel seguro (IPsec) até ao terminal. Além do descrito acima, o ePDG termina os túneis GTP/PMIP com o P-GW como mostra a figura 4.

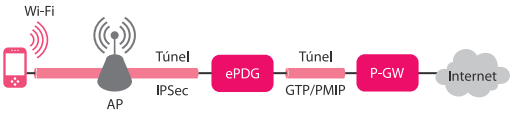


Figura 4. Cenário de acesso non-TWLAN (plano de dados)

É notória a utilização de mecanismos de segurança acrescidos, nomeadamente a utilização de IPSec desde o terminal até ao ePDG, sendo necessário suporte específico no terminal. Embora a maioria dos fabricantes de terminais refiram a intenção de incorporar esta funcionalidade nos dispositivos, não é de desprezar os recursos de processamento que estes requerem.

2.2.b Gestão de Mobilidade

Numa rede heterogénea WLAN + 3GPP, é possível que um terminal esteja ao mesmo tempo na proximidade de várias redes. É assim necessário um controlo de mobilidade e gestão inteligente de rede para uma boa experiência do utilizador e maximização do proveito para o operador, ou seja, balanceamento de tráfego. O 3GPP e o WFA[3] desenvolveram mecanismos e entidades para controlar essa mobilidade, como se explica nas duas secções seguintes.

ANDSF

O *Access Network Discovery and Selection Function* (ANDSF) é introduzido na Release 8 do 3GPP, com o objetivo de ajudar os terminais móveis na descoberta, seleção e ligação a redes de comunicações sem fios, 3GPP ou outras (ex.: WLAN ou *Worldwide Interoperability for Microwave Access* WiMAX). Para além deste elemento servidor na rede do operador, também é necessária a existência de um cliente nos terminais. Entre estas duas entidades foi definido o interface normalizado S14, que se baseia no protocolo *Open Mobile Alliance - Device Management* OMA-DM. Sobre este interface são trocadas políticas na forma de *Management Objects* (MO).

[3] http://www.wi-fi.org/

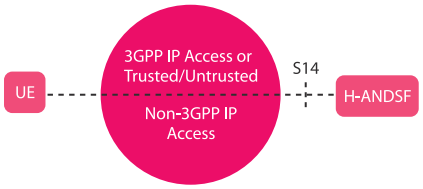


Figura 5. Interface S14 (3GPP/ANDSF)

Essas políticas permitem definir condições de aplicabilidade e prioridades entre redes, para a ligação (*Inter-System Mobility Policy* - ISMP) e para o encaminhamento (*Inter-System Routing Policy* - ISRP) de tráfego. Muito embora o processo de avaliação e execução dessas políticas esteja no domínio do terminal, o operador pode definir quais as regras que se aplicam àquele terminal, numa determinada área geográfica e por um determinado período de tempo. Mediante determinadas condições ou de forma periódica, o terminal contacta o servidor no sentido de receber uma atualização das políticas que lhe são aplicáveis. Também existe forma de o servidor notificar o terminal acerca da existência de novas políticas. Desta forma é possível gerir as ligações dos terminais móveis às redes de um operador.

Na Fig. 6 é visível a evolução de funcionalidades do ANDSF [7]. Esta evolução visa uma melhor gestão de tráfego num ambiente *multi-Radio Access Technology* (RAT), possibilitando informar o terminal sobre qual a rede de acesso mais benéfica para cada aplicação. Na *Release* 12 do 3GPP, o tráfego pode até ser gerido entre *Access Point Name* (APN) diferentes.

Figura 6. Evolução da normalização do 3GPP do ANDSF

HOTSPOT 2.0 e PASSPOINT

O Hotspot 2.0 (HS2.0) é um grupo de trabalho dentro do *Wi-Fi Alliance* (WFA), criado com o objetivo de permitir às redes WLAN um *roaming* tão fácil quanto o atual nas redes 3GPP. O HS2.0 é baseado primariamente nas normas IEEE 802.11u, 802.11i e 802.1x. O HS2.0 Release 1 inclui beacons que os AP enviam. Disponibilizam também uma interface que suporta *Access Network Query Protocol* (ANQP). Entre as informações úteis, é fornecida a carga do AP e os parceiros de *roaming* (lista de *Public Land Mobile Network* (PLMN) que são acessíveis através do AP).

As *features da Release* 1 são complementares às do ANDSF. Já a *Release* 2 compete diretamente com o ANDSF dado que providencia informação de seleção de rede. A recomendação final acordada entre os dois organismos é que dispositivos que possuam as duas tecnologias utilizem a informação proveniente do ANDSF. Complementarmente, o ANDSF está a ser atualizado na *Release* 12 para suportar os itens relevantes do HS2.0 *Release* 2.

2.2.c Segurança

Encriptação *Over-the-air*

As redes 3GPP, como as UMTS e LTE, apresentam interfaces rádio seguros pela adoção de encriptação (128 bits). Numa rede WLAN, a encriptação do interface rádio não é obrigatório, podendo não existir de todo ou ir até níveis iguais aos das redes 3GPP. Para tal é necessário o uso de *Wi-Fi Protected Access 2* (WPA2) que providencia *AES-based airlink encription*. O modo empresarial do WPA2 (WPA2-*Entreprise*) permite fundamentalmente que as chaves sejam dadas por um *Authentication, Authorization and Accounting* (AAA). Para o operador usar ou poder confiar numa rede WLAN como numa rede 3GPP, deve ser usado WPA2-Enterprise.

2.2.d Autenticação e Autorização

A autenticação e autorização em redes 3GPP com núcleo EPC baseiam-se na existência de um *Home Subscriber Server* (HSS). O HSS é uma base de dados que contém informação dos utilizadores e das respetivas subscrições. Adicionalmente, facilita a mobilidade, estabelecimento de sessões e chamadas, sendo também usado pelo núcleo IMS. A autenticação do subscritor em redes EPC é baseada em *Authentication and Key Agreement* (AKA). Este processo permite uma autenticação mútua, garantindo-se que entidades fraudulentas não interferem no processo.

Uma clara dificuldade para a integração das redes WLAN no core EPC consiste na garantia da autenticação do utilizador de uma forma transparente para este. Tradicionalmente, nas redes WLAN, existem 3 formas alternativas de autenticar o utilizador, apresentadas nos parágrafos seguintes:

- Autenticação em portal WEB;
- Autenticação através do uso de certificados;
- Autenticação através das credenciais do (U)SIM.

Autenticação baseada em Portal

Um processo comum de autenticação em redes de acesso WLAN baseia-se em portais *Hypertext Transfer Protocol* (HTTP). Permitem o acesso a utilizadores que ainda não têm uma relação com o operador. Podem aceder com credenciais já detidas ou através da aquisição no próprio portal de, por exemplo, vouchers de tempo ou volume.

Este tipo de autenticação requer a existência de um elemento na rede que faz a interceção de tráfego HTTP, redirecionando o navegador do utilizador para esse portal, bloqueando todo o tráfego enquanto a autenticação não se conclui com sucesso. Normalmente o portal integra-se com um servidor de autenticação *Authentication, Authorization and Accounting* AAA via *Authentication Dial In User Service* (RADIUS).

Autenticação baseada em certificados

Dispositivos sem *Subscriber Identity Module* (SIM) normalmente já suportam a norma 802.1x e podem utilizar qualquer método (por exemplo EAP-TTLS - *Extensible Authentication Protocol-subscriber-Tunneled Transport Layer Security*) que use a autenticação baseada em certificados providenciados pelo operador.

Autenticação através das credenciais do (U)SIM

Dispositivos que possuam um cartão SIM podem usar as credenciais deste para se autenticarem. Existem três métodos de autenticação neste âmbito, para autenticar o cliente em redes WLAN:

- EAP-SIM: baseado no mecanismo EAP e usa as credenciais do cartão SIM.
- EAP-AKA: um melhoramento do método EAP-SIM e usa as chaves simétricas do USIM permitindo uma autenticação entre rede e terminal.
- EAP-AKA': uma revisão do método EAP-AKA que utiliza um mecanismo diferente de derivação das chaves.

De notar que para acesso ao EPC via redes 3GPP baseado na especificação do 3GPP, apenas EAP-AKA e EAP-AKA' devem ser usados.

2.3 QOS EM WLAN

Para permitir ao operador, em redes WLAN, a gestão dinâmica de QoS tal como acontece no resto do EPC, o operador pode usar AP que suportem a norma 802.11u, nomeadamente os mecanismos de acesso ao canal *Enhanced Distributed Channel Access* (EDCA). Mas isto não significa que todos os parâmetros de QoS do EPC tenham um mapeamento direto pois não existem mecanismos de controlo de admissão e de reserva efetiva de recursos como nas redes 3GPP, via um PCRF. Assim, o uso da arquitetura de *Policy Control and Charging* (PCC) não garante um determinado nível de QoS. Permite, contudo, uma melhor engenharia de tráfego na rede WLAN. É disto exemplo a limitação do tráfego instantâneo por utilizador (UE- *Aggregate maximum bit rate* AMBR) e por APN (APN-AMBR). Assim, podem ser aplicadas as mesmas restrições de cotas já aplicadas no EPC. Adicionalmente, as regras de charging da arquitetura de PCC podem igualmente ser aplicadas aos acessos WLAN.

3. MECANISMOS DE *OFFLOADING* E MOBILIDADE

Para a realização dos mecanismos de *offload* e mobilidade de utilizadores entre uma rede de acesso 3GPP e WLAN, existe um conjunto de normalização aplicável, distinguindo-se o seguinte [7]:

- NSWO [8]: *Non-Seamless WLAN Offloading*

Introduzido na *Release* 6 do 3GPP. O UE adquire diferentes endereços IP para as ligações sobre WLAN e 3GPP. Através da ligação WLAN, onde o endereço IP não é alocado pela PGW, o tráfego é encaminhado diretamente para a respetiva rede (Internet), não passando portanto pela PDN-GW no EPC. Neste contexto, não existe integração com o EPC e não existe mobilidade *seamless* (ver Fig. 7).

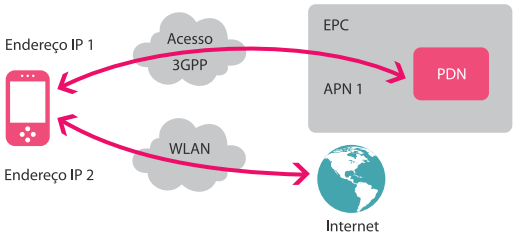


Figura 7. Cenário NSWO

- I-WLAN [4]: *Seamless WLAN Offloading*

A interoperabilidade e mobilidade com WLAN foram introduzidas no EPC na *Release* 8 do 3GPP. Quando um UE está ligado a um acesso WLAN, o tráfego é encaminhado via PDN-GW no EPC. De maneira a suportar a mobilidade e a continuidade da sessão, o UE está ativo apenas num acesso num determinado momento (ver Fig. 8). O mesmo endereço IP é atríbuido pelo PGW, qualquer que seja a rede de acesso.

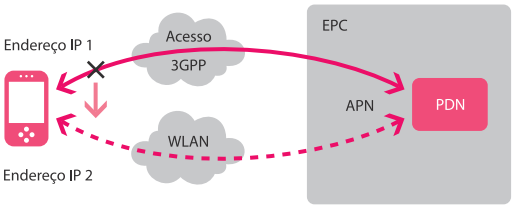


Figura 8. Cenário Seamless WLAN Offloading

- MAPCON [9]: *Multi Access PDN Connectivity*

Introduzido na *Release* 10 do 3GPP, permite ao UE estabelecer simultaneamente múltiplas ligações PDN através de APNs diferentes em redes de acesso distintas (por exemplo 3GPP e WLAN). Permite efetuar o *offload* de algum tráfego de uma rede de acesso para outra, mantendo o mesmo endereço IP (*seamless mobility*). O terminal tem simultaneamente um endereço IP em cada rede de acesso (ver Fig. 10).

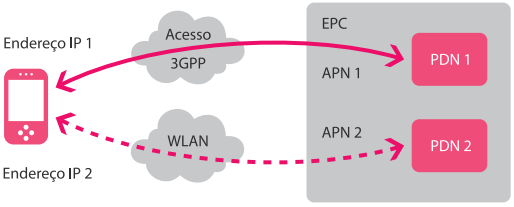


Figura 9. Cenário MAPCON

- IFOM [9]: *IP Flow Mobility and seamless WLAN offload*

O trabalho do IFOM foi introduzido também na *Release* 10 do 3GPP e visa permitir o estabelecimento de fluxos de tráfego simultâneos, sobre múltiplas redes de acesso rádio, partilhando um único endereço IP. Tem como requisito atual a utilização de *Dual Stack Mobile IP* v6 (DSMIPv6), estando a sua implementação com GTP/PMIP em investigação pelo 3GPP.

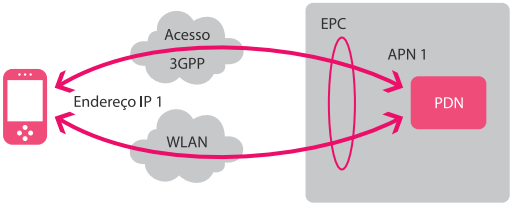


Figura 10. Cenário IFOM

4. IMPACTO EM COMPONENTES PTIN

Como prova este artigo, a integração de WLAN no core EPC está a ser seguida pela PTIN, sendo possível observar o potencial impacto em alguns componentes, nomeadamente:

4.1 PCRF/ANDSF

A interligação entre redes 3GPP e WLAN não requer nenhum suporte em especial por parte da arquitetura de *Policy and Charging Control* (PCC), ou seja, a integração WLAN/3GPP pode acontecer sem alterações no PCRF. Contudo, existem sinergias na utilização de elementos da arquitetura PCC, de modo a ser possível fornecer corretamente convergência e implementar mecanismos de garantia de QoS. O PGW e o PCRF devem adaptar as políticas aplicáveis ao UE consoante a rede de acesso utilizada, e de acordo com o perfil do utilizador. O PCRF, ao ter conhecimento do perfil, permite que o ANDSF crie regras de mobilidade de acordo com o perfil de PCC.

No entanto, no caso da solução IFOM ser implementada numa rede com PCC ativo, o PCRF tem de ser capaz de lidar com múltiplos fluxos para o mesmo IP em múltiplas ligações à rede WLAN/3GPP. O ANDSF por sua vez deve ter definidas políticas ISRP de forma a efetuar uma gestão eficiente do tráfego nos UE.

Estes dois componentes são disponibilizados pela PTIN através da solução ip-Raft.

4.2 AAA

Este componente corresponde ao produto ip-Tiller e necessita ter as funcionalidades necessárias para as interações requeridas para o acesso WLAN e HSS. Para além dos procedimentos de gestão de localização, dos perfis de subscritores e procedimentos de autenticação, tem agora de implementar os interfaces SWx (cenários de *Trusted WLAN*), STa e SWm (cenários de *Untrusted WLAN*).

4.3 ONT-WZG

Relativamente ao acesso WLAN, a solução da PTIN baseia-se atualmente no componente ONT-WZG (*Wireless Zone Gateway*). O WZG integra funcionalidades de um *hotspot*/AP e de um ONT, para interligação a um acesso GPON.

No contexto da integração *trusted* de WLAN no core EPC, este componente deverá suportar a terminação de túneis GRE, assim como a autenticação de terminais via EAP-AKA/AKA'.

5. EXPERIMENTAÇÃO EM AMBIENTE SALINA

Aliando a necessidade da disponibilização de uma montra tecnológica atual e a manutenção de um laboratório transversal à empresa, a plataforma SALINA foi concretizada e é mantida em operação.

Atualmente, na plataforma SALINA está instanciado um core IMS e um core EPC (disponibilizado por um CISCO ASR5000), de maneira a replicar uma rede 4G completa, bem como diferentes acessos WLAN. Na exploração dos vários cenários de integração WLAN, prevê-se a utilização dos vários componentes PTIN, nomeadamente PCRF/ANDSF, AAA, e MyConnect.

6. RECOMENDAÇÕES E CONCLUSÕES

No processo de evolução e convergência em curso, é expectável que a adopção de redes WLAN para complementar a conectividade prestada pelas tecnologias 3GPP seja crescente. Nessa convergência, o núcleo EPC apresenta-se como elemento aglutinador de várias redes de acesso sem fios, 3GPP e não 3GPP.

Adicionalmente perspectivam-se novas formas de conectividade, com os terminais a utilizarem simultaneamente vários interfaces sem fios. Isto apresenta desafios e oportunidades, com impactos em elementos de controlo e plataformas de serviços.

O 3GPP e os principais fabricantes de equipamentos, apresentam soluções com níveis de integração e complexidade diferentes, uns de aplicação imediata, outros com aplicação a médio prazo, quando os terminais possuirem capacidades funcionais e de processamento acrescidas. Neste artigo, sumarizam-se as principais soluções e desafios, segundo vários vectores.

Neste contexto, a PTIN tem produtos que podem integrar soluções de convergência WLAN/EPC, necessitando, no entanto de evoluções de forma a responder aos novos requisitos.

REFERÊNCIAS

[1] M. Olsson, S. Rommer, C. Mulligan, S. Sultana e L. Frid, SAE and the Evolved Packet Core: Driving the Mobile Broadband Revolution, Academic Press, 2009.

[2] 3GPP TS 23.402, "Architecture enhancements for non-3GPP accesses".

[3] RFC 5944, "IP Mobility Support for IPv4," 2010.

[4] 3GPP TS 23.327, "Mobility between 3GPP-Wireless Local Area Network (WLAN) interworking and 3GPP systems".

[5] RFC 5213, "Proxy Mobile IPv6," IETF, 2008.

[6] 3GPP TS 29.060, "General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface".

[7] "Integration of Cellular and WiFi Networks White Paper," 4G Americas, 2013.

[8] 3GPP TS 23.234, "3GPP system to Wireless Local Area Network (WLAN) interworking; System description".

[9] 3GPP TR 23.861, "Multi access PDN connectivity and IP flow mobility".

[10] TS 23.327, "Mobility between 3GPP-Wireless Local Area Network (WLAN) interworking and 3GPP systems," 3GPP Rel-8.

CVS DOS AUTORES

Filipe Rodrigues, concluiu o Mestrado Integrado em Engenharia Eletrónica e Telecomunicações pela universidade de Aveiro. Após o Mestrado, tornou-se bolseiro de investigação do Instituto de Telecomunicações, onde trabalhou em projetos co-financiados pela Fundação para a Ciência e Tecnologia (FCT). Ingressou na PT Inovação no início de 2012 para operacionalizar a plataforma OPEN EPC. Desde então tem participado em desenvolvimentos dos projetos da PTIN para LTE ANDSFnet, CAC LTE e ProbeNI_LTE. Atualmente os seus interesses situam-se na área das comunicações sem fios, rede de acesso e core LTE.

Francisco Fontes, licenciado em Engenharia Electrotécnica e de Computadores, ramo de Electrónica e Telecomunicações, pelo Instituto Superior Técnico (Setembro de 1991) e Doutorado pela Universidade Politécnica de Madrid (Novembro de 2000) na área de gestão distribuída de redes de telecomunicações. Em Setembro de 1991 iniciou a sua actividade profissional na PT Inovação (CET) tendo-se especializado em tecnologias de rede de banda larga. De 2002 a 2012 foi Professor Auxiliar Convidado da Universidade de Aveiro/DETI, para as áreas de redes IP aplicadas às telecomunicações. Actualmente os seus interesses situam-se na área das arquitecturas de redes, em especial na sua evolução para arquitecturas RPG All-IP, com ênfase no IMS. É especialista em tecnologias de rede local e acesso, IPv6 e Multicast.

Nuno Palhares, licenciado em Engenharia Eletrónica e Redes de Computadores pelo Instituto Politécnico de Viana do Castelo (Julho 2010) e com Mestrado em Redes e Serviços de Comunicações da Universidade do Minho (Dezembro 2012). Desde Abril de 2013 encontra-se como estagiário profissional na PT Inovação, estando envolvido na gestão, planeamento e atualização da plataforma SALINA, cobrindo áreas como IMS, EPC, WebRTC.

16 *Software-Defined Networking* Prova de Conceito


BRUNO SENDAS


JOÃO APARÍCIO


JOÃO SOARES


JORGE CARAPINHA


MÁRCIO MELO


RAFAEL GOMES


SUSANA SARGENTO

PALAVRAS CHAVE
Software Defined Networking, Cloud Computing, OpenFlow

O conceito *Software Defined Networking* (SDN) tem ganho um apoio crescente por parte da indústria. Essencialmente, a tecnologia SDN visa tornar as redes programáveis, à medida das características das aplicações e dotá-las de elasticidade e dinamismo, de forma semelhante à computação *Cloud* para o domínio das tecnologias de informação (TI).

Numa fase inicial, o campo de aplicação do SDN foi sobretudo focado nas redes de pequena e média dimensão (e.g. Campus, *Data Centers*), nos desafios colocados pela massificação de tecnologias de virtualização e nos novos requisitos de serviços *Cloud*. O protocolo *Openflow* constituiu nessa fase a peça chave que permitiu demonstrar o conceito e possibilitou a sua rápida popularização. Do ponto de vista dos operadores de telecomunicações, é hoje claro que o SDN tem um campo vasto de aplicações, facilita a inovação e oferece capacidade de diferenciação.

Neste âmbito, tem vindo a ser desenvolvida uma plataforma de rede com a tecnologia *OpenFlow* que permite controlar de forma dinâmica a criação e reconfiguração de várias redes. Esta plataforma é constituída por um conjunto de *OpenFlow switches* e um plano de controlo que detém a inteligência de análise do estado da rede, de decisão e configuração automática de fluxos (*flows*) de dados nos *switches*. Neste artigo serão apresentados alguns casos de estudo experimentais que permitem ilustrar as funcionalidades da plataforma. São ainda discutidas estratégias de migração para as redes dos operadores, concluindo com uma análise sobre a importância desta tecnologia para o grupo PT.

1. INTRODUÇÃO

Recentemente, os requisitos colocados sobre as redes de telecomunicações têm tornado cada vez mais evidentes os seus problemas e limitações. O aumento crescente de dispositivos (sendo grande parte destes dispositivos móveis) e novos paradigmas como o *cloud computing* têm contribuído para exigir da rede requisitos muito mais apertados em termos de dinamismo e flexibilidade. Esta necessidade começou por tornar-se evidente nos domínios *cloud*, i.e. *data centers*, onde existem processos contínuos de alocação e reconfiguração de recursos virtuais, exigindo adaptações constantes por parte da rede. No entanto, o impacto não está limitado às redes de *data center*, na medida em que a rede de transporte desempenha igualmente um papel fundamental quer na interligação de *data centers*, quer na entrega dos serviços ao cliente final. É portanto necessário que as redes atuais evoluam para uma arquitetura mais ágil, capaz de se adaptar aos novos requisitos.

É neste sentido que surge o paradigma de *Software Defined Networking* (SDN), que propõe uma mudança na arquitetura da rede, com a separação entre os planos de controlo e de dados. Do ponto de vista lógico, esta separação permite colocar a inteligência da rede num plano superior, reduzindo assim a complexidade dos elementos de rede e tornando a rede mais programável. O método de operação assenta na troca de mensagens entre o plano de controlo e o plano de dados, passando a rede a ser um instrumento dinâmico, em contraste com o modelo tradicional, baseado na configuração estática dos elementos de rede. Neste momento, o *Openflow* é o protocolo de referência para efectuar a comunicação entre os dois planos.

É importante realçar que não se está a falar de um conceito meramente teórico, pois já existe uma oferta significativa de produtos no mercado. Entre outros, podem mencionar-se *NEC*, *Nicira* (*Vmware*), *Cisco* e *Alcatel-Lucent* como fabricantes de equipamentos de rede com suporte para *OpenFlow* (ou outros protocolos proprietários), e também de *software* de controlo para redes SDN. Existem também já vários casos de implementação em grande escala, sendo o caso da Google um dos mais significativos, na medida em que utiliza uma rede *OpenFlow* para fazer a gestão de tráfego entre os seus diferentes *data center* [1]. Outros *cloud vendors* já seguiram os mesmos passos, como é exemplo da *CloudSigma* [2] que utiliza SDN na gestão da rede interna dos seus *data centers*. No entanto, as implementações não se cingem apenas a fornecedores de *Cloud*, tendo a *Nippon Telegraph and Telephone* (NTT) sido, tanto quanto sabemos, o primeiro operador de telecomunicações a implementar SDN não apenas em *data center*, mas também em redes de transporte. Este caso concreto voltará a ser referido mais à frente.

Este artigo pretende dar a conhecer o conceito de SDN e as suas potencialidades, o que de mais relevante tem sido feito nesta área, bem como o impacto que poderá ter para o grupo PT. A parte restante do presente artigo apresenta a seguinte estrutura: a secção 2 introduz o conceito SDN bem como o protocolo *OpenFlow*; na secção 3 é apresentada uma plataforma experimental SDN que pretende demonstrar algumas das potencialidades de SDN e do *OpenFlow*; a secção 4 olha para o SDN numa perspetiva da sua aplicação em ambientes *Telco*. Alguns dos possíveis impactos da tecnologia no grupo PT são identificados na secção 5. Finalmente, a secção 6 apresenta as conclusões e considerações finais, bem como o trabalho futuro.

2. *SOFTWARE-DEFINED NETWORKING* E *OPENFLOW*

SDN é uma tecnologia emergente que apresenta uma nova visão para a conceção, desenvolvimento e exploração das infra-estruturas de rede, baseada no princípio da separação dos planos de controlo e de dados, permitindo assim o desenvolvimento de uma entidade lógica central, a partir da qual pode ser efetuada a configuração e gestão dos elementos de rede, e oferecendo uma maior flexibilidade, abstração e visão global da rede, com o objetivo de ter um suporte universal, isto é, independente do *hardware* [3] [5].

A arquitectura SDN é constituída por três planos, como podemos observar na Figura 1.

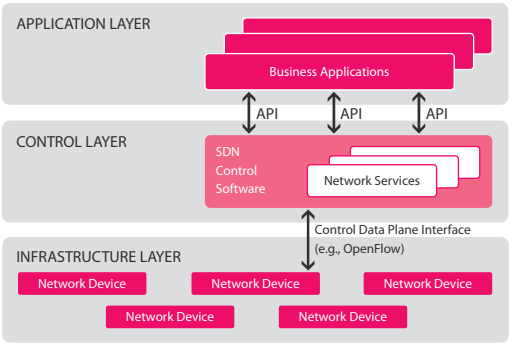


Figura 1. Planos da arquitetura SDN [6]

O plano inferior da arquitetura SDN diz respeito aos dispositos de rede e é denominado de plano de infra-estrutura (*infrastructure layer*), ou plano de dados. Apresenta uma interface que permite a comunicação com o plano imediatamente superior.

O plano de controlo (*control layer*) é responsável por configurar, gerir e mapear os serviços solicitados pelo plano de aplicação (*application layer*) na rede física. Este plano encontra-se desvinculado do plano de infra-estrutura, proporcionando assim uma visão geral e centralizada da rede como um todo e não como elementos independentes interligados.

Por último, o plano da aplicação é responsável por efetuar a tradução dos requisitos das aplicações (largura de banda, latência, segurança e qualidade de serviço (QoS)) em aplicações de rede [6].

O protocolo mais aceite no âmbito SDN é o *Openflow* [5]. A sua função básica é a comunicação entre os planos de infra-estrutura e de controlo. O *Openflow* requer um *software* de controlo externo, responsável por controlar o encaminhamento da informação de um dispositivo de rede. A principal abstração introduzida na especificação *Openflow* é o conceito de fluxo (*flow*). Entende-se por *flow* o conjunto de campos do cabeçalho, ou seja, as *flags* ativas previamente configuradas que identificam o tipo de pacote a ser processado pelo dispositivo de rede. Por análise das *flags*, o dispositivo define o encaminhamento a seguir pelo pacote na rede, assim como as acções futuras a realizar para pacotes semelhantes que cheguem à rede. A especificação atual do protocolo (versão 1.4) garante o suporte da camada 2 a 4 do modelo OSI, existindo também suporte ao nível da camada 1 para redes óticas. No entanto, o grupo responsável por esta especificação está a trabalhar no suporte das camadas 4 a 7.

O plano de dados de um dispositivo de rede é constituído por uma tabela de regras/encaminhamento e filtros configuráveis. A cada *flow* que entra num dispositivo são comparados os campos do cabeçalho do pacote com a informação presente na tabela de encaminhamento. O resultado da comparação poderá resultar no descarte ou envio do pacote, ou eventualmente no envio com modificação de um ou mais campos.

O princípio de funcionamento do *Openflow*, ilustrado na Figura 2, é o seguinte:

1. Um pacote chega a um dispositivo de rede com suporte para o protocolo de comunicação *Openflow* (*switch Openflow*).

2. Seguidamente, o *switch* compara os campos do cabeçalho do pacote com a sua tabela de regras, a fim de saber se tem alguma configuração para esse tipo de pacote, podendo-se verificar uma de duas situações:
   - No caso de não existir um *flow* para o tipo de pacote, o switch envia o cabeçalho do pacote para o controlador, ficando este responsável pela tomada de decisão sobre a ação a realizar no mesmo. Neste caso, o controlador pode optar pelas seguintes operações:
     - Descarte do pacote, sendo lançado um evento com a informação '*miss table*', que sinaliza ao controlador e ao *switch* que, para esse tipo de pacote, não apresenta regra que defina o tráfego;
     - Inserção de uma regra de acordo com as características do pacote que permita efetuar o seu encaminhamento. Seguidamente, o controlador envia novamente o pacote para o *switch* em que se ativou esse tipo de *flow*.
   - No caso de existir a configuração de *flow* para o pacote em questão e se verificar o *match* com a tabela de regras, o pacote é encaminha-

>>>

do de acordo com a configuração incluída no seu cabeçalho.

3. Seguidamente são actualizados os contadores dos *switches* e realizadas as acções pré-configuradas.

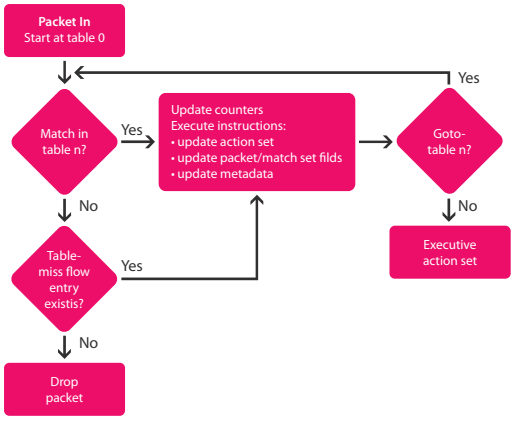


Figura 2. Funcionamento do protocolo *Openflow* [9]

Entre as vantagens que a tecnologia SDN e o *OpenFlow* proporcionam podem evidenciar-se:

- Independência do *hardware*: a deslocação da inteligência da rede para um plano superior permite criar uma abstração do plano inferior, ou seja, dos elementos individuais presentes na rede [6].

- Redução da complexidade: a deslocação do plano de controlo para um plano superior (ou seja, a passagem da inteligência dos elementos de rede para um controlador [6]) permite que a infra-estrutura de rede passe a ter elementos muito mais simples, que apenas necessitam de executar funções de transporte.

- Facilidade e Inovação: as SDN permitem que haja uma segmentação das redes em planos. Com esta segmentação é possível ter sob o substrato diferentes redes independentes, permitindo assim o teste de tecnologias disruptivas sem prejudicar o normal funcionamento das outras redes [6].

- Dinamismo: a separação do plano responsável pela configuração da rede e do plano de elementos de rede permite uma programação dinâmica do tráfego de rede [6].

ESTADO DA ARTE

Actualmente, existe uma forte aposta da indústria em SDN e também no protocolo *OpenFlow*. São vários os fabricantes com ofertas comerciais de equipamentos de rede *OpenFlow*. Foi esta a aposta inicial dos fabricantes, onde podemos salientar: *NEC, HP, Pica8, Juniper, Brocade, Arista Networks* e *Ericsson* [7]. Mais recentemente a aposta tem sido nas plataformas de controlo SDN, como: a solução *Contrail* da *Juniper*; *Trema* da *NEC*; *Open Network Environment* (ONE) da *Cisco*; *Nuage* da *Alcatel-Lucent*; e *Big Network Controller* da *Big Switch Networks*.

Enquanto ao nível dos equipamentos de rede a adoção do protocolo *OpenFlow* é praticamente unânime, o mesmo não acontece ao nível da camada de controlo da arquitectura SDN. No sentido de, entre outras coisas, colmatar esta lacuna, surge no início de 2013 o projeto *OpenDaylight* [8]. Este projecto nasce de uma aliança industrial, onde podemos encontrar, entre outros: *Cisco, Ericsson, Juniper, Citrix, Redhat, Intel, Plexi, NEC, Vmware*. É importante realçar que a adesão ao consórcio é aberta a qualquer empresa que queira contribuir. De igual importância é o facto de que todos os desenvolvimentos são tornados públicos, tanto especificações como *software*.

3. PLATAFORMA EXPERIMENTAL SDN

A plataforma SDN apresentada nesta secção permite a criação e gestão de serviços de conetividade sobre uma rede *Openflow*. Além de assegurar a integridade desses serviços (gestão de falhas), também gere de uma forma ótima o uso da rede (gestão em tempo real).

Para a plataforma, um serviço de conetividade é definido por quatro parâmetros: tipo de conetividade, tipo de serviço, terminais e tipo de regra. O primeiro parâmetro define o tipo de comunicação, por exemplo, comunicação *Multicast* ou *Full-Mesh*. O segundo define o tipo de serviço de rede, por exemplo, comunicação TCP (*Transmission Control Protocol*), UDP (*User Datagram Protocol*), ICMP (*Internet Control Message Protocol*) ou ARP (*Address Resolution Protocol*). O terceiro define o conjunto de terminais que fazem parte do serviço, sendo estes identificados através de endereços MAC e/ou IP. Em alternativa à identificação de terminais, pode ser identificada a porta do switch. O último parâmetro é o tipo de regra a aplicar na rede, por exemplo, identificando o IP (*Internet Protocol*) e/ou MAC (*Media Access Control*) de origem ou destino presente no cabeçalho dos pacotes. Parâmetros de QoS não foram tidos em conta devido a limitações no controlador que apenas suporta a versão 1.0 do *Openflow*.

A arquitetura da plataforma, presente na Figura 3, é composta por cinco módulos, quatro dos quais compõem o plano de controlo da estrutura: *Flow*

Handler, *Activator*, *Monitoring* e *Controller*. O plano de controlo liga o plano de dados, onde estão os elementos de encaminhamento, ao plano de aplicação, onde reside o quinto módulo – *Service Handler*. O plano de controlo é ainda responsável por assegurar a otimização do funcionamento da rede. As funções individuais dos diferentes módulos são descritas de seguida.

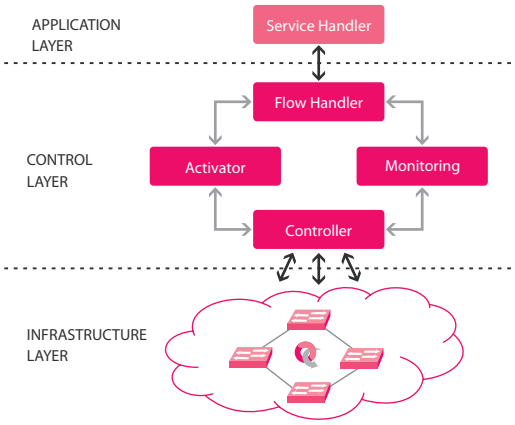


Figura 3. Arquitetura da Framework SDN

- O ***Service Handler*** é o ponto de entrada da *framework*, permitindo aos utilizadores efetuar pedidos de instanciação de serviços. Suporta três funcionalidades: o *Service to Flow Translation*, *Service Repository* e *Flow Repository*. O primeiro é necessário para fazer a tradução de serviços em *flows*, uma vez que o ***Flow Handler*** não tem a visão de serviços. Os seguintes permitem guardar a informação necessária para uma descrição completa de serviços e fluxos.

- O ***Flow Handler*** pode ser visto como o cérebro da rede, sendo responsável por encontrar os melhores caminhos segundo os algoritmos definidos. Este módulo possui três funcionalidades: *Flow Allocation, Flow Optimization* e *Fault Detection*. Os dois primeiros estão relacionados com a descoberta de caminhos, sendo aplicados em diferentes contextos. O *Flow Allocation* é o processo responsável por mapear *flows* aquando da receção de novos pedidos de serviços de conetividade. O *Flow Optimization* é responsável por re-otimizar um serviço previamente instanciado, quando necessário. A última funcionalidade deste módulo diz respeito à identificação de potenciais problemas na rede, sendo feita para isso uma comparação da topologia de rede actual com a anterior.

- O módulo ***Monitoring*** é o elemento agregador de toda a informação da rede e é composto por três funcionalidades. A função de *Real Time Network Discovery* permite construir uma representação da topologia da rede e das características das ligações entre os elementos. As funções de *Network Statistics* e *Event Generation* encontram-se interligadas. A primeira gera informação estatística da utilização da rede que irá desencadear o envio de mensagens de notificação e de alarme através da segunda funcionalidade.

- O módulo ***Activator*** baseia-se em duas funcionalidades, *Flow* Translation e *Flow* Enforcement, utilizadas de modo sequencial para configurar as tabelas de encaminhamento dos switches. A primeira faz a tradução de *flows* em regras separadas por switches e a segunda consiste na comunição com o controlador para a aplicação das mesmas.

- Por último, o ***Controller*** é o elemento responsável pela abstração dos elementos de rede para os módulos nas camadas superiores. Para isso irá utilizar a função de *Network Device Mediation* para estabelecer a ponte entre a *framework* e os elementos de rede. Neste caso recorremos ao *Floodlight* [10], um controlador SDN aberto suportado pela *Big Switch Networks* (o *Floodlight* é parte integrante da solução comercial *Big Network Controller da Big Switch Networks*). A escolha recaiu sobre o *Floodlight* devido à sua constante evolução e suporte da comunidade, bem como pelo facto de disponibilizar uma *API RESTful*, facilitando assim a separação deste dos outros elementos.

A Figura 4 apresenta a interação entre os vários módulos para diferentes momentos: inicialização do sistema, ativação de serviço, monitoria, deteção de falha e processo de re-otimização. Ao longo do processo, verifica-se uma comunicação contínua entre os módulos *Monitoring* e *Controller* mediante recurso a uma *API RESTful* e também entre o *Controller* e a rede, com recurso ao protocolo *Openflow*.

Quando o sistema é iniciado, o módulo *Flow Handler* subscreve os serviços expostos pelo *Monitoring*. Simultaneamente, o *Monitoring* recolhe do *Controller* a informação dos elementos de rede (e.g. número de portas, largura de banda das ligações), que por sua vez a fornecerá ao *Flow Handler* após processamento. Terminada esta fase, a plataforma encontra-se disponível para receber e ativar os serviços de conetividade.

Na ativação de um serviço, o utilizador comunica ao *Service Handler* o serviço que pretende ativar na rede. Este procede à validação do pedido e ao seu armazenamento numa base de dados. O *Flow Handler* en-

>>>

contra-se em comunicação contínua com a base de dados à procura de novos *flows* para ativar. Quando este módulo encontra novos *flows* na base de dados, obtém a sua descrição e procura um caminho na rede. Para o processo de procura de caminho recorremos ao algoritmo de *Dijsktra* [11] (que toma em consideração a largura de banda das ligações no substrato para escolher o melhor caminho). Por fim, transmite todas as informações necessárias ao módulo *Activator* para ativar os *flows* na rede. A ativação dos *flows* por parte do módulo *Activator* é uma atualização das tabelas de regras de cada um dos elementos da rede que fazem parte do caminho seleccionado, que é efetuada através da interação com o *Controller*.

O processo de re-otimização é iniciado quando o módulo *Monitoring* deteta uma alteração da informação da rede obtida pelo *Controller*. Este módulo notifica o *Flow Handler*, módulo subscritor, fornecendo-lhe as alterações verificadas no substrato. Por comparação com a informação disponível anteriormente, o *Flow Handler* analisa se foram adicionados ou removidos elementos ou ligações na rede. No caso de adição, este irá procurar por possíveis otimizações aos serviços ativos. No caso de eliminação de um elemento ou ligação, este irá procurar possíveis falhas nos serviços ativos e verificar se existe uma outra forma de reativar o serviço (i.e. os seus respetivos *flows*).

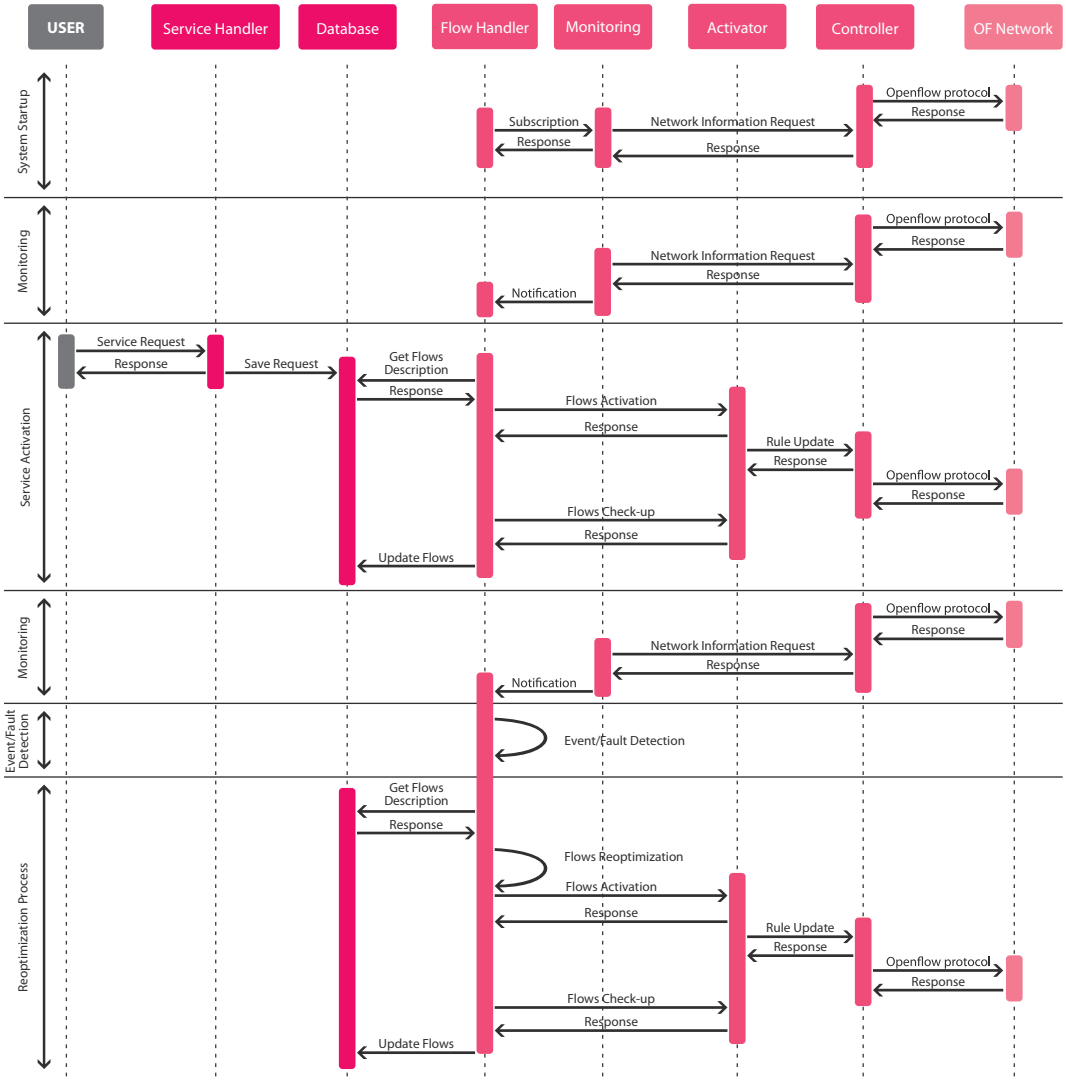


Figura 4. Interação entre módulos (*System Startup, Service Activation* e *Reoptimization Process*)

CASOS DE ESTUDO

A fim de analisar as funcionalidades da plataforma foram estudados 3 casos:

1. **Ativação de serviço**: Efetua-se o pedido de ativação de um serviço à plataforma; esta recebe e avalia a consistência do mesmo. O pedido é traduzido em *flows* individuais, tendo em conta o tipo de serviço a ativar, armazenando-os na base de dados. De seguida, é executado um algoritmo de procura de caminhos para cada um dos *flows*. Por fim, os *flows* são ativados na infra-estrutura. Este caso de estudo compreende toda a fase *Service Activation* presente no diagrama da Figura 4.
   Exemplo de pedido de serviço: tipo de serviço - UDP; tipo de comunicação - *Multicast*; elemento origem - nó A; elementos destino - 2 no nó B e 2 no nó D. Resulta num pedido com 5 elementos, que vai ser traduzido em 4 *flows*.

2. **Perda de *link***: A plataforma deteta a perda de um *link* ao verificar uma alteração da topologia de rede. São identificados os *flows* afetados e de seguida é desencadeado o respectivo processo de re-otimização. Por fim, é feita a reconfiguração dos elementos da infra-estrutura a fim de manter os serviços ativos. Neste caso, são abrangidas 3 das fases presentes no diagrama da Figura 4: *Monitoring*, *Event/Fault Detection* e *Reoptimization Process*.

3. **Adição de *link***: De forma semelhante à perda de *link*, a plataforma deteta a presença de um novo *link* ao verificar uma alteração na topologia de rede. Esta alteração desencadeia uma re-optimização completa de todos os *flows*, de acordo com o novo estado da rede. De forma semelhante ao caso anterior, as fases abrangidas neste caso são: *Monitoring*, *Event/Fault Detection* e *Reoptimization Process*.

É de salientar que as políticas utilizadas nesta prova de conceito têm um intuito experimental, não tendo que corresponder exatamente às melhores políticas a adotar numa rede real.

A Figura 5 apresenta a testbed desenvolvida onde foram postos em prática os casos de estudo. Os elementos de rede são simples desktops a correr instâncias *Open vSwitch* [12] com suporte para a versão 1.0 do protocolo *Openflow*.

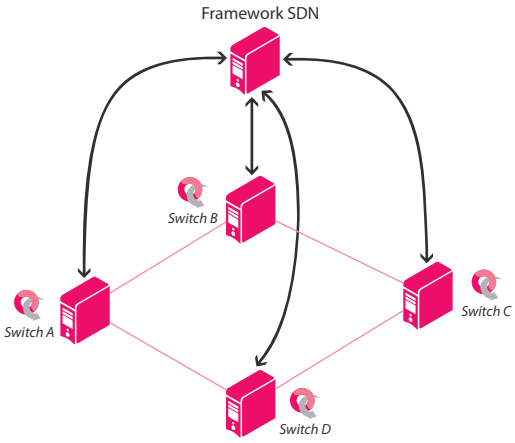
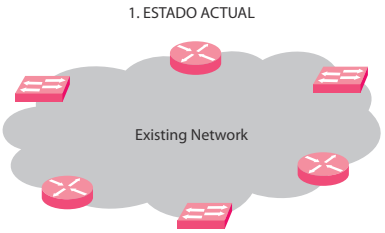


Figura 5. *Testbed*

4. SDN EM AMBIENTES *TELCO*

Do ponto de vista dos operadores de telecomunicações é hoje claro que a tecnologia SDN e o protocolo *OpenFlow* apresentam um vasto campo de aplicações, facilitam a inovação e oferecem capacidade de diferenciação. É portanto expectável que nos próximos anos venhamos a assistir a um aumento na oferta de serviços com base na tecnologia *OpenFlow*. No entanto, será necessário cumprir um conjunto de requisitos em termos de escalabilidade, desempenho, fiabilidade e robustez exigidos numa rede de operador, o que envolve alguns desafios. Além disso, será importante definir um modelo que permita a coexistência das redes *OpenFlow* com as redes tradicionais, tendo em conta as características disruptivas desta tecnologia.

Atualmente, alguns operadores já exploram soluções baseadas em SDN[1]. Para tal, é necessário integrar uma solução SDN com a infra-estrutura de rede tradicional do operador. Este processo de integração terá que prever também a possível migração faseada de uma rede híbrida (*Openflow* e equipamento tradicional) para uma rede totalmente *OpenFlow* [14]. A Figura 6 ilustra este processo.



[1] O operador de telecomunicações japonês NTT é um exemplo de sucesso desta integração, tendo hoje uma oferta de *cloud computing* com um conjunto de funcionalidades inovadoras. Através da inclusão de SDN, os clientes podem realizar configurações nas redes em tempo real, incluindo ainda a possibilidade de ajustar parâmetros de QoS segundo um modelo de negócio *on-demand* [9].

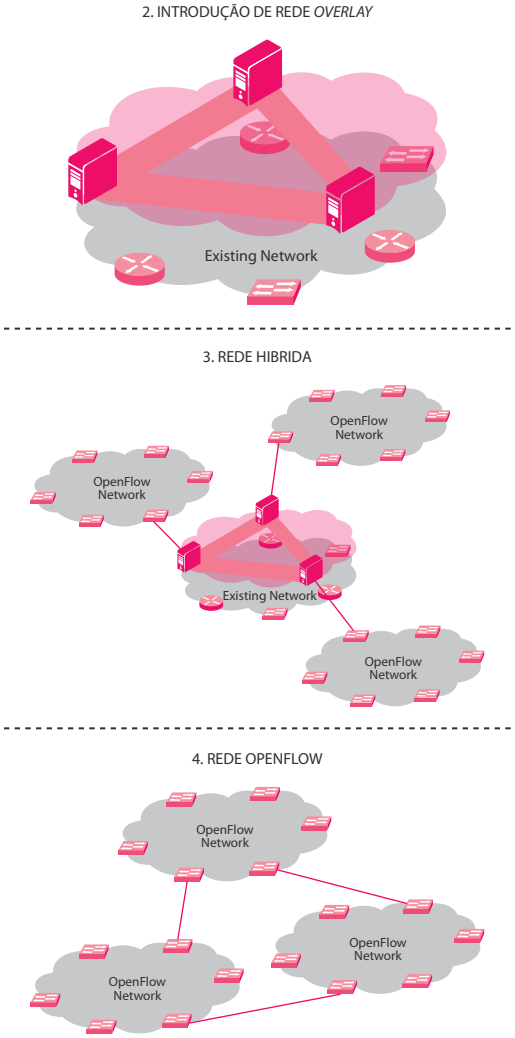


Figura 6. Processo de integração de SDN na rede atual [14]

Para que este modelo possa ser introduzido numa rede de operador tradicional, é necessário em primeiro lugar garantir o isolamento e uma boa integração entre as redes *OpenFlow* e a restante infra-estrutura. Garantido o cumprimento deste requisito, pode iniciar-se o processo de aprovisionamento da infra-estrutura que irá assegurar os novos serviços baseados em *OpenFlow*. Na Figura 6, a infra-estrutura de rede tradicional do operador é responsável por ligar os *switches OpenFlow* entre si isolando-os do restante tráfego na rede, representado na figura pelos túneis que interligam os switches. Do ponto de vista do equipamento *OpenFlow* existe uma única infra-estrutura de rede, abstraindo a rede tradicional que apenas serve de transporte entre ilhas *OpenFlow*.

Recorrendo a este modelo é possível a coexistência das redes tradicionais de operador com redes compatíveis com o protocolo *OpenFlow*. Este modelo permite ainda fazer uma migração entre uma infra-estrutura tradicional e uma infra-estrutura assente em tecnologia SDN. Em qualquer momento é possível expandir a rede *OpenFlow*, sendo apenas necessário fazer a ligação do novo equipamento com o túnel apropriado, ou ligando diretamente a outro equipamento *OpenFlow* já existente[2].

5. IMPORTÂNCIA PARA OS NEGÓCIOS DO GRUPO PT

As secções anteriores deram uma panorâmica do potencial da tecnologia SDN para o operador, tanto em termos de oportunidades como de desafios. É evidente que estamos perante uma nova mudança de paradigma nas tecnologias de rede, que poderá ter um impacto relevante no negócio dos operadores. Embora, numa fase inicial, a implementação da tecnologia se tenha concretizado sobretudo em ambientes de *data center* (*Cloud*), a sua implementação tende a alargar-se para ambientes do operador de telecomunicações.

É portanto fundamental para o grupo PT não só reter conhecimento tecnológico na área, como também perceber os possíveis impactos que a tecnologia poderá implicar. Os possíveis impactos destas transformações tecnológicas no negócio do grupo PT podem ser avaliados segundo duas vertentes – por um lado, enquanto operador e fornecedor de serviços de rede; por outro lado, como potencial fornecedor de sistemas de rede e OSS/BSS:

- Como detentor de uma rede de cobertura nacional, deverão ser avaliados os benefícios em termos de gestão interna da infra-estrutura pela aplicação da tecnologia SDN.

- Como fornecedor de serviços *Cloud*, deverá ser avaliada a hipótese de aplicar a tecnologia nos *data center* PT, bem como na entrega dos serviços ao cliente. A oferta combinada de serviços *Cloud* e *Telco* foi já discutida em dois artigos anteriores desta revista - "*Clouds* de próxima geração" [16] e "Gestão Integrada de Serviços *Cloud* e *Telco*" [17].

- Como fornecedor de serviços tipicamente denominados *Telco*, deve considerar-se a possibilidade de, usando tecnologias SDN, alargar o leque de

[2] Neste âmbito, deve mencionar-se a recente apresentação do projecto japonês O3 (O Three) Project [12], que junta a NEC, NTT, Fujitsu e Hitachi. O projecto pretende criar a primeira *Wide Area Network* (WAN) completamente baseada em SDN, e prevê, entre outras coisas, uma redução potencial de 90% no tempo de planeamento e construção de WANs.

ofertas nesta área e dotar a rede dos atributos de elasticidade e automatismo que permita a oferta de serviços "on-demand".

- Como fornecedor de equipamentos de rede, deverá ser avaliada a hipótese de dotar os equipamentos com o protocolo *OpenFlow*.

- Como fornecedor de Sistemas de Suporte à Operação (OSS – *Operation Support Systems*), deverão ser avaliados os impactos da arquitectura SDN na operação e gestão da rede e a hipótese de dotar este tipo de sistemas com a capacidade de interagir com redes *OpenFlow*.

6. CONCLUSÃO E TRABALHO FUTURO

Os requisitos que são hoje colocados sobre as redes de telecomunicações têm vindo a evidenciar a necessidade de evoluir para novos modelos e arquitecturas, de forma a conseguir dar a resposta adequada a novos requisitos de agilidade e automação. É neste contexto que o conceito do SDN tem vindo a ganhar um apoio crescente por parte da indústria, incluindo os operadores. Apesar de ainda haver um longo caminho a percorrer até estas tecnologias poderem ser consideradas maduras para uma adoção em larga escala, parece não haver dúvidas de que a evolução para arquitecturas de rede privilegiando a agilidade e a programabilidade é uma tendência irreversível.

Este artigo pretendeu dar a conhecer melhor o conceito e os fatores diferenciadores do SDN, bem como o protocolo *OpenFlow* como a maior referência neste domínio. Foi ainda apresentada uma plataforma experimental SDN recorrendo ao protocolo *OpenFlow* como prova de conceito de algumas das funcionalidades inerentes à tecnologia. A forma como o SDN e as redes atuais podem coexistir foi também abordada neste artigo, bem como as possíveis áreas de impacto da tecnologia no grupo PT.

Como trabalho futuro pretendemos avaliar algumas das mais proeminentes soluções disponíveis (ou que irão estar disponíveis em breve), nomeadamente a primeira stack de arquitectura e *software* que irá ser disponibilizada pelo projecto *OpenDaylight* [7]. O *OpenContrail* [16] (*stack* de *software* da plataforma SDN *Contrail* da *Juniper*) é outra possível plataforma alvo. A utilização de plataformas normalizadas e interfaces abertas tenderá a facilitar o desenvolvimento de aplicações que permitirão tirar partido do SDN, de forma efectiva.

Outro objectivo para trabalho futuro é a integração de SDN com a virtualização de funções de rede, que constitui outra tendência de evolução muito relevante nesta área, com um impacto expectável igualmente significativo para os operadores.

REFERÊNCIAS

[1] Google, "Inter-Datacenter WAN with centralized TE using SDN and OpenFlow", 2012, URL: https://www.opennetworking.org/images/stories/downloads/sdn-resources/customer-case-studies/cs-googlesdn.pdf.

[2] Jason Verge: "CloudSigma Version 2.0 Embraces SDN and SSDs", URL: http://www.datacenterknowledge.com/archives/2013/06/20/cloudsigma-unveils-2-0-the-qualitative-cloud/?utm-source=feedburner&utm-medium=feed&utm-campaign=Feed%3A+DataCenterKnowledge+%28Data+Center+Knowledge%29. Consultado a 24 Setembro 2013.

[3] M. Mendonca, B. Nunes, X.Nguyen, K. Obraczka and T. Turletti: "A Survey of Software-Defined Networking: Past, Present and Future of Programmable Networks", May, 2013.

[4] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner, "OpenFlow: enabling innovation in campus networks."ACM SIGCOMM Computer Communication Review 38.2 (March 2008): 69-74.

[5] W. Stallings: "Software-Defined Networks and Openflow", in The Internet Protocol Journal, Cisco, Volume 16, Issue 1, pp. 2-14, March, 2013.

[6] Open Networking Foundation: "Software-Defined Networking: The New Norm for Networks", white paper, April, 2012.

[7] SDN Product Directory, http://sdndirectory.opennetworking.org/products. Consultado a 24 Setembro 2013.

[8] OpenDaylight Project, http://www.opendaylight.org/. Consultado a 24 Setembro 2013.

[9] Open Networking Foundation: "Openflow Switch Specification", August, 2013.

[10] Project Floodlight, http://www.projectfloodlight.org/. Consultado a 24 Setembro 2013.

[11] E. W. Dijkstra: "A note on two problems in connexion with graphs", Numerische Mathematik 1, pp. 269–271, 1959.

[12] Open vSiwtch, http://openvswitch.org/

[13] NTT, NTT Communications to Launch World's-first SDN-based Cloud Migration Service, NTT Communications, URL: http://www.ntt.com/aboutus_e/news/data/20130627.html. Consultado a 24 Setembro 2013.

[14] NTT, NTT DATA's Advance in SDN Business Provides Highly-Flexible Control of Network by Software, NTT DATA Corporation, URL: http://www.nttdata.com/global/en/news-center/global/2012/060801.html. Consultado a 24 Setembro 2013.

[15] O3 Project launched for achieving the world's first wide area SDN, NEC, URL: http://www.nec.com/en/press/201309/global_20130917_01.html. Consultado a 24 Setembro 2013.

[16] João Soares, Romeu Monteiro, Márcio Melo, Jorge Carapinha, Pedro Neves e Susana Sargento: "Clouds de Próxima Geração", Saber & Fazer Telecomunicações, 2011.

[17] Pedro Neves, João Soares, André Augusto, Mário Rui Costa e Alexandre Laranjeira: "Gestão Integrada de Serviços Cloud e Telco", Saber & Fazer Telecomunicações, 2012.

[18] OpenContrail Project, http://opencontrail.org/. Consultado a 19 de Outubro de 2013.

CVS DOS AUTORES

Bruno Sendas, M.Sc. Integrado em Engenharia Electrónica e Telecomunicações pela Universidade de Aveiro em 2012. Iniciou a sua actividade profissional no Instituto de Telecomunicações, pólo de Aveiro, através de uma bolsa de investigação científica em Maio do mesmo ano. Participou como colaborador na fase final do projeto SAIL – *Scalable and Adaptive Internet Solutions*, na componente de *Cloud Networking*, sendo neste momento colaborador do projecto *Mobile Cloud Networking*. Os seus principais interesses focam-se na interligação de rede com *cloud*: *cloud computing, cloud networking, software defined-networking* e *network functions virtualization*.

João Aparício, M.Sc Integrado em Engenharia de Electrónica e Telecomunicações pela Universidade de Aveiro em 2013. Iniciou a sua actividade profissional no Instituto de Telecomunicações, pólo de Aveiro, através de uma bolsa de investigação iniciada em Setembro do mesmo ano. As suas áreas de interesse englobam gestão de recursos, *cloud computing/networking*, redes móveis e *software-defined networking*.

João Soares, M.Sc. Integrado em Engenharia de Electrónica e Telecomunicações pela Universidade de Aveiro em 2009. Iniciou a sua atividade profissional no Instituto de Telecomunicações, pólo de Aveiro, através de uma bolsa de investigação científica em Setembro do mesmo ano. Atualmente frequenta o programa doutoral em Engenharia Electrotécnica na Universidade de Aveiro, tendo em Outubro de 2010 obtido uma Bolsa de Doutoramento em Empresa co"nanciada pela FCT e pela PT Inovação como estagiário no departamento de Coordenação Tecnológica e Inovação Exploratória. Foi colaborador ativo do projeto SAIL – *Scalable and Adaptive Internet Solutions*, na componente de *Cloud Networking*, sendo neste momento colaborador do projeto *Mobile Cloud Networking*. Os seus interesses contemplam mobilidade, gestão de redes e qualidade de serviço, bem como as temáticas da cloud: cloud computing, mobile *cloud computing*, *cloud networking* e *software defined-networking*.

Jorge Carapinha, obteve a Licenciatura em Engenharia Electrotécnica, Ramo de Informática, pela Faculdade de Ciências e Tecnologia da Universidade de Coimbra (1984) e Mestrado em Electrónica e Telecomunicações pela Universidade de Aveiro (1998). Desde 1985 é colaborador da PT Inovação (anteriormente CET). No âmbito de projetos nacionais e internacionais tem desenvolvido atividades em diversos domínios, com destaque para tecnologias de redes IP/MPLS, redes privadas virtuais e qualidade de serviço. O seu trabalho mais recente tem estado focado nas áreas de *Cloud Networking*, Redes Definidas por *Software* e Virtualização de Funções de Rede.

Márcio Melo, Doutorando em Engenharia Eletrotécnica com M.Sc. Integrado em Engenharia Eletrónica e Telecomunicações pela Universidade de Aveiro (2008). Iniciou a sua atividade profissional no Instituto de Telecomunicações em Setembro de 2008, no polo de Aveiro, através de uma bolsa de investigação científica. Em Dezembro de 2009, obteve uma Bolsa de Doutoramento em Empresa cofinanciada pela FCT e pela PT Inovação. Ingressou nesse ano na PT Inovação como estagiário no departamento de Investigação Aplicada e Difusão do Conhecimento onde participou ativamente em projetos europeus do 7º Programa-Quadro (FP7) como o 4WARD – "Future Internet" e SAIL – "Scalable and Adaptive Internet Solutions". Colaborou ainda num estudo Eurescom: "Network Virtualisation – Opportunities and Challenges for Operators". Atualmente encontra-se inserido no departamento de Desenvolvimento de Sistemas de Rede, a evoluir a linha de produtos NetBand da PT Inovação, mais especificamente o segmento de acesso "GPON-IN-A-BOX". Membro do IEEE Graduate Student nº: 90624478. As suas áreas de interesse passam por Redes de Acesso Ótico Passivo de Próxima Geração (NG-PON), Redes Metro Ethernet MPLS-TP (CE2.0) e Redes Core IP/MPLS, bem como temas de investigação que versem a Virtualização de Rede ou o *Cloud Networking*.

Rafael Gomes, M.Sc Integrado em Engenharia de Electrónica e Telecomunicações pela Universidade de Aveiro em 2013. Iniciou a sua actividade profissional no Instituto de Telecomunicações, pólo de Aveiro, através de uma bolsa de investigação iniciada em Setembro do presente ano. As suas áreas de interesse enclobam redes de telecomunicações, gestão de redes, *software defined network, cloud computing*.

Susana Sargento, (http://www.av.it.pt/ssargento) obteve o Doutoramento em 2003 em Engenharia Electrotécnica na Universidade de Aveiro. Ela foi docente do Departamento de Ciências de Computadores da Universidade do Porto de Setembro de 2002 a Fevereiro de 2004, e encontra-se na Universidade de Aveiro e no Instituto de Telecomunicações desde Fevereiro de 2004, onde actualmente lidera o grupo de Arquitecturas e Protocolos de Redes (http://nap.av.it.pt). Ela faz parte também do corpo docente convidado do Departamento de Eng. Electrotécnica e de Computadores da Universidade de Carnegie Mellon, USA, desde Agosto de 2008, onde realizou 'faculty exchange' em 2010/2011. Desde Março de 2012, a Susana é co-fundadora de uma empresa de redes veiculares, a Veniam'Works, que tem como objectivo construir uma rede Internet com base em veículos. A Susana tem estado envolvida em vários projectos nacionais e internacionais, tendo liderado algumas actividades, como as actividades de qualidade de serviço e de redes ad-hoc do projecto europeu FP6 IST-Daidalos. Ela esteve recentemente envolvida em vários projectos FP7 (4WARD, Euro-NF, C-Cast, WIP, Daidalos, C-Mobile), projectos nacionais, e projectos do programa CMU|Portugal (DRIVE-IN com a Universidade de Carnegie Mellon). Os seus interesses de investigação centram-se nas áreas de redes de nova geração e redes da Internet do futuro, mais especificamente em QoS, mobilidade, redes auto-geridas e cognitivas. A Susana faz regularmente parte do Painel de Especialistas nos programas de investigação europeus.

17 Arquitetura de Rede FTTH para Oi


ANDERSON LEAL


ANTÓNIO VIDAL


FABRÍCIO CASTRO


FERNANDO MORGADO


JOÃO CACHUCHO


JOÃO FIGUEIREDO


KLEBER CAMARA


MIGUEL DIZ


RUI MORAIS

PALAVRAS CHAVE
FTTx, Topologias de Rede de Acesso, Infraestruturas de Rede, Materiais de Rede Passivos, Projeto de Rede

A Oi como maior empresa de rede fixa no Brasil decidiu seguir uma estratégia de implementação de Rede de Nova Geração em fibra ótica baseada na tecnologia GPON. Nesse sentido, desenvolveu um estudo teórico com o objetivo de adquirir conhecimento em redes Fiber To The x (FTTx) que permitisse desenvolver soluções de arquitetura e topologias a implementar.

Tendo definido uma base de trabalho, era altura de viabilizar soluções, e escolheu a PT Inovação como empresa parceira a nível técnico devido ao conhecimento e experiência em redes de fibra ótica. A empresa foi ao encontro do cliente de forma a obter um enquadramento do trabalho desenvolvido, bem como da rede de infraestrutura existente, o meio, os processos, entre outros, para que, de forma clara e objetiva, passar à apresentação de soluções no que refere à arquitetura de rede e processos a implementar.

Este artigo pretende evidenciar o quão necessário é ouvir o cliente, identificar o que realmente necessita e auxiliá-lo a atingir a qualidade desejada. Pretende-se de igual forma fundamentar as principais soluções técnicas, bem como os processos adotados na implantação da rede da Oi.

1. INTRODUÇÃO

O mercado residencial tem vindo a exigir cada vez mais largura de banda de forma a suportar novos serviços, tais como, Internet Protocol television (IPTV), Video on Demand (VoD), High Speed Internet (HSI), entre outros. Estes serviços têm tido uma procura e adesão crescente a nível mundial, tornando-se de importância vital para as operadoras. No Brasil, segundo números da Anatel e a título de exemplo, o serviço de televisão por assinatura nos últimos anos conta com um crescimento médio anual de cerca de 30%.

As redes FTTx surgem pelo facto de os serviços a oferecer aos potenciais clientes exigirem suportes com capacidades de transmissão crescentes. O desenvolvimento deste tipo de redes tem no entanto que ser cuidadosamente planeado devido ao seu elevado custo, tendo sempre em conta a sua flexibilidade e escalabilidade. Várias topologias são apontadas, sendo que todas partem do princípio de que se deve minimizar o número de equipamentos a instalar. A ocupação de espaços é também relevante, não só ao nível das instalações do operador, como nas condutas e câmaras de visita associadas à infraestrutura de suporte. Por todas estas razões, as soluções apontadas para este tipo de redes convergem sempre em topologias ponto-multiponto com uso de elementos de divisão de potência ótica ("splitters") passivos. De acordo com os casos em concreto pode optar-se por terminar a rede ótica num armário de rua (Fiber To The Node - FTTN ou Fiber To The Cabinet - FTTC), num armário na base dos edifícios (Fiber To The Building - FTTB) ou dentro da habitação dos clientes (Fiber To The Home - FTTH).

Prever arquiteturas de rede convergentes com forte interação por parte do cliente, onde um conjunto de serviços é disponibilizado sobre múltiplos acessos/clientes, já é uma realidade. O aumento de tráfego associado a novos hábitos de consumo, com o acréscimo de novos serviços e a natural pressão criada pela concorrência, poderá provocar dinâmicas não previstas.

O investimento necessário à construção de uma rede de acesso é de tal forma elevado que o objetivo passa por construir uma rede eficiente e robusta, à prova do futuro, de baixo CAPEX/OPEX, prevendo um crescimento sustentável, provocando uma enorme pressão aquando do desenvolvimento e definição de uma rede desta natureza.

A Oi definiu arquiteturas de rede numa abordagem onde as centrais ditavam a capacidade máxima, independentemente da concentração de unidades de cliente existentes no raio de ação das mesmas. Na sequência das várias reuniões, onde foram apresentadas e justificadas abordagens diversas que têm na sua essência as melhores práticas em redes FTTx aplicadas às necessidades do cliente, a Oi solicitou a colaboração da PT Inovação na definição da melhor arquitetura de rede a implementar.

Em seguida será apresentada a abordagem inicial desenvolvida pela Oi e consequentemente a abordagem final, resultado das muitas interações entre a PT Inovação e a Oi.

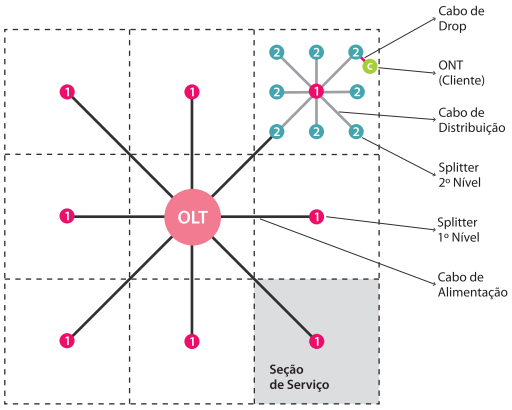
ABORDAGEM INICIAL OI

Numa primeira fase, a Oi decidiu iniciar o desenvolvimento de topologias apenas para projeto FTTH pelo elevado potencial de suporte a serviços de banda larga com o maior grau de escalabilidade no universo das redes FTTx. Uma rede com um investimento inicial mais elevado contudo permite

>>>

economias no que se refere a custos de manutenção, grande parte devido à inexistência de equipamentos ativos entre a central e as instalações do cliente. O objetivo passou também por iniciar a construção em áreas nobres no que concerne ao padrão de clientes, bem como à qualidade das infraestruturas de suporte. Em áreas mais antigas das cidades onde as infraestruturas de suporte são naturalmente mais antigas e já não suportam mais cabos nos edifícios, condutas e caixas de visita, a Oi previu o desenvolvimento de topologias para projeto de redes FTTB e FTTC ou FTTN.

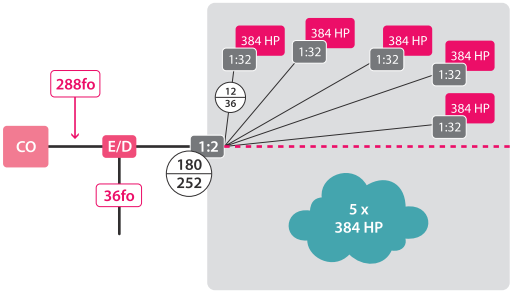
A primeira abordagem às redes FTTH tratava-se de uma visão *top-down* onde a *central office* era a prioridade. Às centrais era conferida uma capacidade máxima e uma configuração de cabos típica, fazendo-se depois uma distribuição da capacidade inicialmente prevista à volta, conforme representado na figura seguinte:



Central office e respetivos cabos óticos

Era definido um raio de cobertura de acordo com o alcance teórico de uma PON (Passive Optical Network) e dividida esta área em oito secções de serviço mais a secção central. Da central para o cliente, a rede era constituída pela rede primária (cabo de alimentação), rede secundária (cabo de distribuição) e rede drop, designada muitas vezes como rede de cliente. Sairiam oito cabos de 288 fibras óticas com o intuito de servir cada uma das oito secções de serviço mais um oitavo da secção central. A rede secundária era toda ela constituída por cabos de 36 fibras óticas.

Entre a rede primária e secundária realizar-se-ia o primeiro andar de splitting e um segundo andar de splitting a realizar-se entre as redes secundária e drop. A figura abaixo representa o esquema de cabos e splitters para dar atendimento a uma determinada secção de serviço:



Distribuição dos andares de splitting na rede

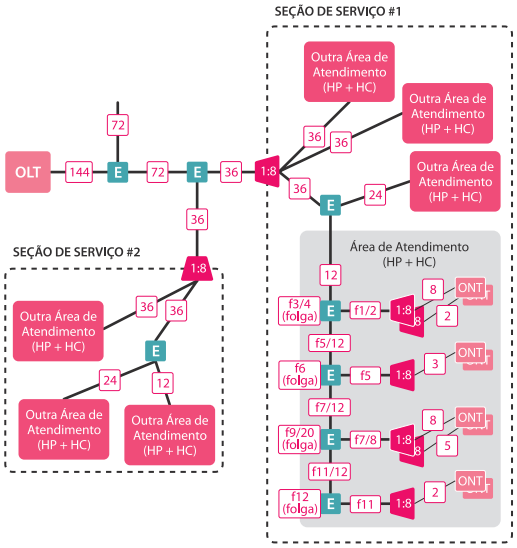
Três topologias foram previstas para implementação da rede, cada uma delas com diferentes combinações de splitters no primeiro e segundo níveis. A tabela seguinte resume as combinações de splitters às três topologias definidas. O tamanho da célula e do ramo dependem, neste caso, pela capacidade do splitter de segundo nível associado à capacidade do cabo de rede secundária, 36 fibras óticas.

Cada secção de serviço serviria até 3840 unidades de cliente, pelo que a central office (estação telefónica) teria uma capacidade de serviço até 34560 (3840 x 9 secções) unidades de cliente.

Topologia	Combinação de Splitters	Célula	Ramo	Secção de serviço	Estação telefónica
A	1:4 + 1:16	192	960	3.840	34.560
B	1:2 + 1:32	384	1920	3.840	34.560
C	1:8 + 1:8	96	480	3.840	34.560

Topologias de rede

Analisando a figura abaixo podemos verificar que na maioria dos casos torna-se difícil atingir uma combinação de splitters de acordo com as necessidade de cada aglomerado de ONTs. O resultado é o natural desaproveitamento não só de toda a infraestrutura de rede mas também dos equipamentos ativos na central, usualmente conhecido por portos ociosos.

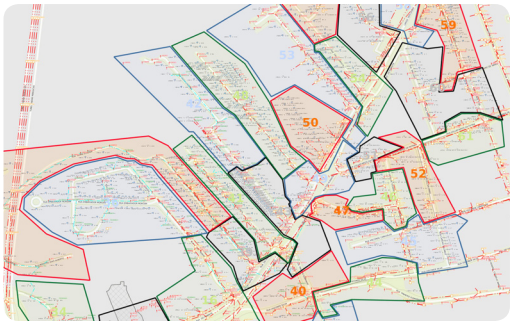


Rede de distribuição ótica - Topologia C

ABORDAGEM FINAL

Com o avançar do processo foi possível à Oi confrontar o seu plano inicial com a experiência da PT Inovação na área da temática GPON e rever a abordagem no sentido de a tornar mais eficiente e de acordo com a realidade existente no terreno. Definiu-se assim um conjunto de regras para avançar com o projeto FTTH:

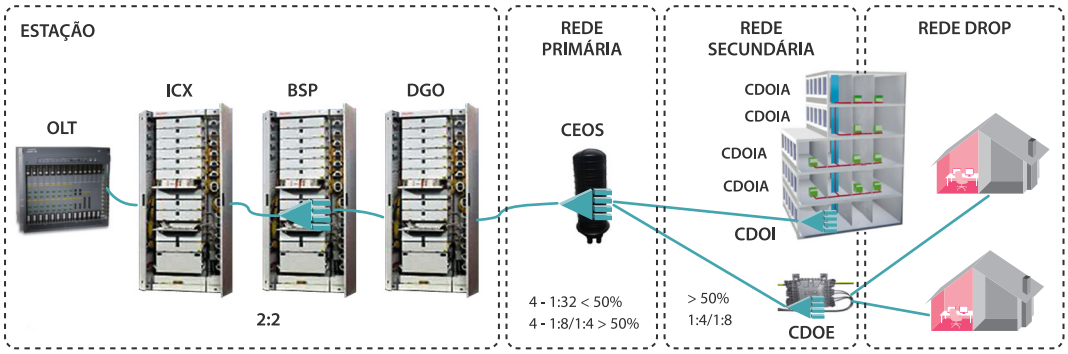
- O departamento comercial define as áreas a atender e as respetivas prioridades.
- É realizado o site survey com base na informação da comercial e infraestruturas de suporte existentes, segundo necessidades do projeto de engenharia.
- Todo o processo é iniciado do cliente para a central, trata-se de uma visão down-top onde o atendimento a cada cliente é prioridade.
- A engenharia agrupa os clientes em células e garante rede para até 133% de atendimento nas áreas definidas pelo departamento comercial. O processo é denominado por celularização e consiste na divisão da área de influência de uma central em células. Todo o cliente de uma célula é atendido a partir de uma Caixa Distribuição Ótica (CDO), por sua vez é alimentado pela rede secundária que converge numa caixa de emenda ótica com *splitters* (CEOS) ou num Armário Rua Distribuição Ótica (ARDO). A figura abaixo é exemplo do resultado dos processos de celularização.



Celularização

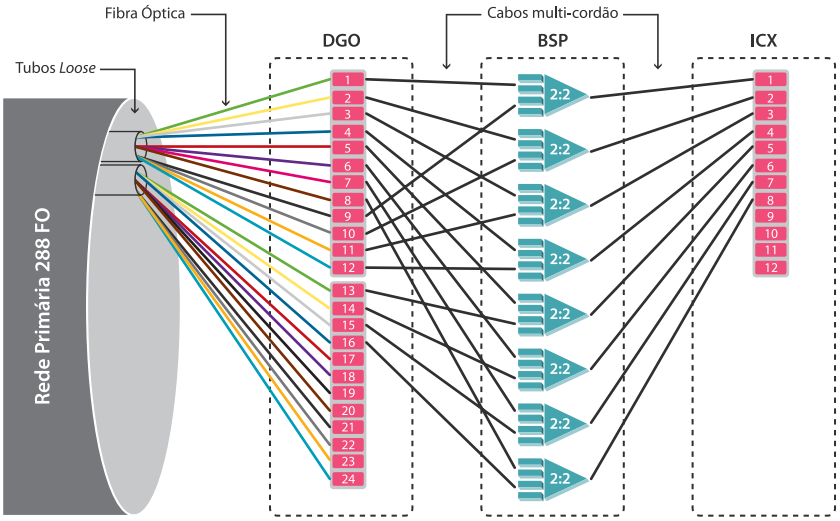
- É garantida uma rede *future-proof* limitando a rede OSP para um rácio de splitting máximo 1:32. O rácio de 1:64 é feito na Central Office com o uso de *splitters* do 2:2.
- A engenharia introduziu um ponto de flexibilidade na rede com um *coupler* 2:2 na central. Do primeiro nível de *splitting* 2:2 a realizar-se na central resultam as vantagens:
  - Criação do primeiro ponto de flexibilidade na central;
  - Simplificação no aumento de débito por cliente e alcance geográfico da rede.
- Foram definidas diferentes fases de atendimento garantindo na primeira fase de instalação até 50% dos clientes. O atendimento acima de 50% é realizado pontualmente nas caixas de distribuição ótica através da instalação de splitters, passando parte do CAPEX para OPEX com a intenção de resultar numa solução *pay-as-you-grow*. Em todos as caixas de distribuição é feito um escalonamento com fases de instalação que está dependente da venda ou seja, instala-se numa primeira fase aproximadamente 50% do numero de clientes, atingido este numero vão ser instalados *splitters* até se obter os 100%.

Tipo de Arquitectura	Descrição de Etapas			Tipo de Ligação de 2º Andar de Splitting	Critérios de Aplicação	
	1º Etapa de Splitting (Central)	2º Etapa de Splitting (ARDO/ CEOS/CDO)	3º Etapa Splitting (CDO)			
A	2:2	1:32	--	Splitter pre-conectorizado (entrada e saída) colocado na CEOS	Para conectar até 50% do total de UC's do edificio	*A Oi definirá inicialmente que areás ou células de devem desenhar com topologia A ou B (dependerá da possibilidade de colocar armários na planta exterior, ou do orçamento de potência). * Serão áreas de edificios (distribuição interior ou exterior) que tenham pontos de distribuição ou colunas montantes com menos de 64 UC's. * Também será aplicado em casos em que temos um conjunto de edificios (distribuição interior) cuja soma de UC's seja inferior ou igual a 64 UC's e que tenham um único ponto de acesso comum a todos eles, considera--se todo o conjunto como um único edificio com um nº total de UC's igual a soma dos UC's de cada.
	2:2	1:4	1:4		Para conectar de 50% a 100% do total de UC's do edificio (UC's <= 24 UC's)	
	2:2	1:4	1:8		Para conectar clientes de 50% a 100% do total de UC's do edificio (UC's >24)	
B	2:2	1:32	--	Splitter fusionado (entrada e saída) colocado no armário (ARDO)	Para conectar até 50% do total de UC's do edificio	
	2:2	1:8	1:4		Para conectar de 50% a 100% do total de UC's do edificio (UC's <= 24 UC's)	
	2:2	1:4	1:8		Para conectar de 50% a 100% do total de UC's do edificio (UC's >24)	
C	2:2	1:32	--	Splitter pre-conectorizado na sua saida colocado no interior do (CDOI)	*Para edificios (distribuição interior) que tenham um ponto de distribuição com mais de 64 UC's. *Para casos em que temos um conjunto de edificios (distribuição interior) cuja soma de UC's seja igual ou superior a 64 UC's e que tenham um unico ponto de acesso a todos eles , considerando todo o conjunto como um unico edificio com um nº total de UC's igual a soma das UC's a cada edificio do conjunto.	



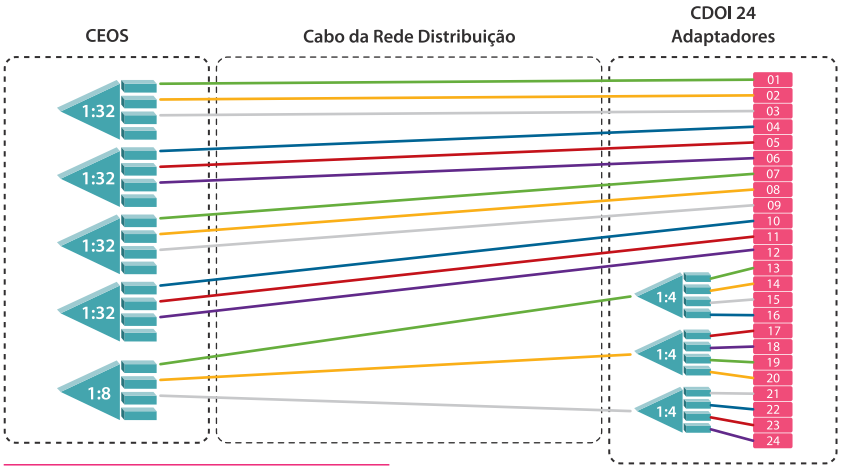
Fases de atendimento

- Foi introduzido o conceito de Emalhamento tanto na rede de distribuição quer na rede primária, com isto a Oi passou a conseguir praticamente anular os portos ociosos, garantir o balanceamento da rede ou seja evitar "apagões" e fazer uma redução muito grande de numero de potos da OLT.



Emalhamento na central office

>>>



Emalhamento na caixa de emenda ótica com splitters

A figura acima é exemplo do emalhamento realizado para atendimento de uma determinada caixa de emenda ótica a atender 24 unidades de cliente. Numa primeira fase, o atendimento até 50% é realizado através de *splitters* 1:32 na CEOS e a segunda fase, acima de 50%, o atendimento é realizado através da instalação de *splitters* 1:4 no CDO. Esta garante não só a possibilidade de realizar um crescimento pontual e gradual mediante a necessidade como também a eliminação de portos ociosos.

- O estudo das fibras de reserva é feito em função da topologia de construção dos cabos de fibra ótica. O resultado é a diminuição do número de fusões, respetiva diminuição no budget ótico (orçamento de potência) e facilidade na construção da rede.

- Algumas das soluções acima mencionadas foram desenvolvidas com o intuito de evitar ao máximo futura sobreposição de redes. As figuras abaixo são meramente representativas do padrão das redes de acesso em cobre da Oi.

Rede de cabos aéreos na rede de acesso

Pormenor da rede de cabos em poste

Foi também feito um esforço de normalização na documentação de todo o processo, simbologia e apresentação dos entregáveis de forma que diferentes equipas apresentem documentação coerente e facilmente interpretável.

Simultaneamente correram os processos de caraterização dos materiais, com registo de informação através do desenvolvimento de manuais, ações de formação e consultoria. Sempre que necessário a Oi conta com o apoio da PT Inovação para desenvolvimento e definição de melhorias, respetivas atualizações de manuais e formação contínua.

2. CONCLUSÃO

A Oi adotou como estratégia a renovação da sua rede de acesso, optando por investir numa rede FTTx que permitisse capacidades de transmissão muito elevadas. Iniciou o processo com um estudo teórico que comparou vantagens e desvantagens das várias soluções FTTH, FTTC, FTTB e FTTN e de diferentes topologias de rede. No entanto, a abordagem inicial da Oi apresentava algumas características passíveis de serem otimizadas, nomeadamente na mitigação de portos ociosos. Trabalhando em conjunto com a PT Inovação foi possível estudar novas abordagens que permitiram criar uma solução de rede com caraterísticas de *pay-as-you-grow*, *future-proof* e *hardware optimization*.

REFERÊNCIAS

[1] www.anatel.gov.br
[2] www.teleco.com.br
[3] www.ftthcouncil.org
[4] ETT 027-2011 - Rede FTTx - Requisitos de Engenharia
[5] Manual Site Survey Oi, Rio de Janeiro, Brasil, 2012
[6] Dossier de Projeto de Rede FTTH Oi, Rio de Janeiro, Brasil, 2012
[7] Manual de Especificação de Materiais FTTH Oi, Rio de Janeiro, Brasil, 2012
[8] Manual Projeto de Rede e Cadastro FTTH Oi – Netwin, Rio de Janeiro, Brasil, 2012

CVS DOS AUTORES

Anderson Leal, graduado em Engenharia Elétrica pela Universidade Gama Filho (2000), com especializações em Eletrônica e Telecomunicações, e pós-graduado em Engenharia Electrotécnica e de Computadores pelo Instituto Superior Técnico da Universidade de Lisboa (2004), na área de Telecomunicações. Ingressou na PT Inovação Brasil em 2005, na equipe Netband, onde forneceu serviços e equipamentos de transmissão, qualidade de serviço, agregação e otimização de tráfego, para a rede móvel. Desde 2012 é o responsável pelo Projeto de Fornecimento de Treinamento, Manuais e Fiscalização, na Construção de Rede Externa FTTH na Oi.

António Vidal concluiu, em 1988, o Curso Técnico Profissional de Eletrónica – Área de Estudos "B" – Científico Tecnológicos Na Escola Secundária Nº1 de Aveiro. Ingressou no CET – Centro de Estudos de Telecomunicações no ano 1991 onde assumiu tarefas no CETLAB – Laboratório de Equipamentos Terminais. Desde 2011 faz parte da PT Inovação, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede na área de redes FTTx.

Fabrício Peixoto de Castro, graduado em Engenharia Elétrica pelo Centro Federal de Educação Tecnológica do Rio de Janeiro (2007) e em Tecnologia de Rede de Computadores pela Associação Brasileira de Ensino Universitário (1998). Ingressou na Oi em 1998, atuando em diversas áreas como Operação e Manutenção de rede metálica; Na construção de rede óptica, coordenando a célula de mão de obra própria teve atuação marcante no PAM (Plano de Antecipação de Metas) que garantiu a Oi exploração antecipada da Telefonia Móvel; Na gestão da implantação de redes ópticas do Rio de Janeiro conduziu a implantação de todo a infraestrutura óptica para a realização dos jogos Pan-americanos de 2007 e para a grande expansão da banda larga, Projeto Velox, que resultou na Implantação de 1 milhão de portas Velox (ADSL2+) em 1 ano; Na diretoria de Banda Larga, atuou como PMO no projeto UBL (Ultra Banda Larga) preparando a rede Oi para oferta de velocidades de banda larga acima de 10MB; Desde 2011 atua no planejamento de rede óptica, fazendo parte da equipe responsável por todas as definições para as redes FTTx da Oi.

Fernando Morgado, licenciado em Engenharia Electrónica e Telecomunicações pela Universidade de Aveiro em 1984. Mestrado em Telecomunicações pela Queen Mary University of London em 2006. Ingresso no Centro de Estudos de Telecomunicações dos CTT, actual PT Inovação, em 1988 onde efectuou trabalho de desenvolvimento em sistemas de transmissão por fibra óptica durante os primeiros anos de actividade profissional. Posteriormente exerceu a sua actividade em vários projectos relacionados com tecnologias e arquitecturas de rede de acesso, no âmbito dos programas ACTS e IST da Comunidade Europeia, bem como do Eurescom. Participou ainda em pilotos tecnológicos ADSL, VDSL, APON, transmissão óptica por técnica de solitões e WDM. Exerceu funções ligadas ao laboratório de Testes de Certificação e Conformidade de equipamentos de telecomunicações na PT Inovação. Atualmente exerce funções de formação e certificação de pessoas no âmbito de equipamentos desenvolvidos na PT Inovação.

João Cachucho concluiu, em 2009, licenciado em Engenharia Civil pelo Instituto Superior de Engenharia do Porto. Iniciou-se a colaborar para a PT Inovação, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede, no início de 2009 no âmbito do projeto de cadastro de infraestruturas de rede. Integra desde 2011 a equipa de serviços de projeto, cadastro e consultoria de redes FTTx.

João Figueiredo concluiu, em 1995, Superior Technician in Electronics na Universidade de Reading U.K. Ingressou em 2001 na PT Inovação, nos Serviços de Engenharia e desde 2005 que é responsável pela unidade de desenvolvimento e produção de infraestruturas e serviços de FTTx, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede.

Kleber de Oliveira Camara formado em Administração de Empresa (2008) – Curso de Gestão de Projeto pela Compass/Dinsmore Associates 2012/2013 – Trabalha na OI desde 1998, passou pelas áreas de operação, engenharia e planejamento da OI, onde participou de vários projetos importantes como PAM (Plano de Antecipação de Metas) que garantiu a OI explorar a Telefonia Móvel; ProjetoTriunfo – grande expansão e melhoria da rede da OI com a criação de novas estações prediais; Projeto Velox – Implantação de 1 milhão de portas Velox (ADSL2+) em 1 ano; projeto UBL (Ultra Banda Larga) preparando a rede OI para oferta de velocidades de banda larga acima de 10MB e Projeto FTTH/GPON OI onde atuou desde o nascimento até a implantação dos primeiros clientes.

Miguel Diz, concluiu, em 2008, a Licenciatura em Engenharia Eletrotécnica e de Computadores na Faculdade de Ciências e Tecnologia da Universidade de Coimbra. No mesmo ano ingressou na PT Comunicações através do programa Trainees para a área de operação e manutenção da rede de acesso GPON/xDSL. Desde 2011 faz parte da PT Inovação, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede. No inicio de 2012 assumiu funções de Team Leader da equipa de serviços de projeto, cadastro e consultoria de redes FTTx.

Rui Morais concluiu, em 2007, o Bacharelato em Engenharia Eletromecânica e em 2008, a Licenciatura em Engenharia Eletrotécnica (ramo Mecatrónico). Colabora desde 2008 na PT Inovação, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede, onde acompanhou o arranque dos projetos de cadastro de redes e projeto de redes de fibra ótica (FTTx). Integra atualmente o grupo de desenvolvimento de infra estruturas de redes de telecomunicações.

18 Otimização de CAPEX em Redes FTTx

JOÃO FIGUEIREDO MIGUEL DIZ

PALAVRAS CHAVE
FTTx, Infraestruturas de Rede, CAPEX, OPEX, *Pay-as-You-Grow, Future-Proof, Unbundling*, Rede de Acesso

O mercado das telecomunicações caracteriza-se por estar em constante evolução. Para se manterem competitivas, as operadoras de telecomunicações têm de ser inovadoras e adotar tecnologias que lhes permitam fornecer aos seus clientes serviços diferenciadores face à concorrência. Muitas dessas tecnologias requerem suporte em infraestruturas de rede de alto débito, obtidas através da implementação de redes FTTx (*Fiber-To-The-x*).

A implementação de novas redes de acesso pressupõe um grande investimento, sendo que o CAPEX (*Capital Expenditure*) anual de algumas operadoras chega a aumentar 30% com a construção destas redes de nova geração. Considerando ainda o retorno lento do investimento, torna-se fundamental que as operadoras façam um planeamento cuidado por forma a minimizarem o risco financeiro.

Este artigo pretende decompor o investimento realizado pelas operadoras na implementação de redes de acesso FTTx, apresentar fatores importantes a considerar no planeamento inicial para além do CAPEX, analisar as atividades inerentes ao processo de construção destas redes e verificar onde e como reduzir os custos. Por fim, apresentar-se-ão alguns exemplos de otimização de CAPEX na seleção dos materiais passivos e no desenho do projeto de rede.

1. INTRODUÇÃO

Nos últimos anos, operadores de todo o mundo têm apostado em renovar as suas redes de acesso, construindo infraestruturas de rede em fibra ótica FTTx. De facto, segundo dados do FTTH Council (*Fiber-To-The-Home Council*), no final de 2009 existiam na Europa (EU35) cerca de 16 milhões de casas ligadas a redes FTTx passando a haver 34 milhões de casas em 2012. Ou seja, houve um crescimento anual (*CAGR – Compound Annual Growth Rate*) de aproximadamente 31% em HP's (*Home Passed*). De forma análoga, verificou-se um crescimento anual de 39% para o número de subscritores, sendo que atualmente já existem mais de 7,3 milhões de clientes ligados a redes FTTx (ver Figura 1).

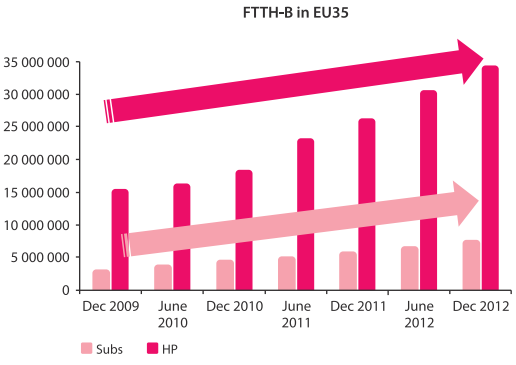


Figura 1. Evolução do nº de subscritores e casas passadas de 2009 a 2012 (dados FTTH Council)

A razão para os operadores estarem a investir em infraestruturas de rede de acesso em fibra ótica prende-se com o facto de estas possibilitarem o uso de tecnologias de transmissão de ultra banda larga. Esta característica permite fornecer novos serviços que infraestruturas legadas não permitem. Serviços como HDTV (*High-definition Television*) (que requer cerca de 20 Mbit/s por canal), telemedicina, jogos *online*, entre outros. Muitos operadores têm adotado esta estratégia por representar uma oportunidade de ganharem cota de mercado e consequentemente aumentarem as receitas. Contudo, a implementação de uma nova rede de acesso implica investimentos avultados. Para minimizar e proteger esse investimento é necessário fazer um planeamento rigoroso, analisando todas as componentes do custo inicial.

Na Figura 2 pode-se ver a decomposição do custo inicial num cenário sem infraestruturas existentes (*greenfield*). Cerca de 46% do investimento inicial é gasto com trabalho civil (abrir valas, instalação condutas, lançamento de cabo, instalação de elementos de rede,…) sendo a componente com maior potencial de redução de custo. De seguida, verifica-se que 26% do investimento necessário é gasto com os equipamentos ativos nas centrais e 16% com a instalação no cliente. Por fim, apenas 12% do custo é para materiais (cabos, juntas de fibra ótica, armários, …). Esta análise permite identificar a importância dos operadores terem uma infraestrutura de rede pré-existente com condutas ou postes.

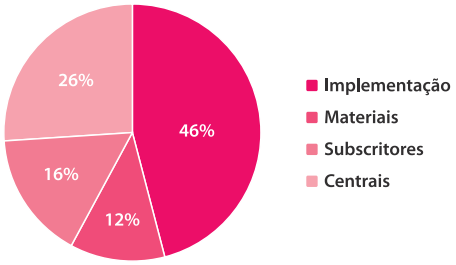


Figura 2. Decomposição do investimento inicial num cenário *greenfield* (dados FTTH Council)

2. REDUZIR VS. OTIMIZAR CAPEX

Ao planear a construção de uma rede FTTx as operadoras devem ter em conta vários fatores. Muitas abordagens focam-se em reduzir CAPEX e descartam, por exemplo, os gastos anuais de operação e manutenção da rede (OPEX) (*Operational Expenditure*). Considerando a previsão de que as novas infraestruturas de rede em fibra ótica fiquem em exploração pelo menos durante 30 anos, rapidamente se entende a importância de incluir o OPEX como fator decisivo. O facto da rede de cobre em Portugal ter sido instalada há mais de 5 décadas e ainda hoje continuar a ser explorada pelos operadores através de tecnologias como o VDSL (*Very-high-bit-rate Digital Subscriber Line)* suporta esta teoria. Se com um investimento ligeiramente superior for possível reduzir o OPEX, então é a decisão certa a longo prazo até porque o retorno do investimento (*ROI*) é previsivelmente lento. Além disso, com OPEX reduzidos o *break-even* acontecerá mais cedo.

É também fundamental tomar decisões que não comprometam o *upgrade* futuro de novas tecnologias de transmissão, isto é, tem de se garantir a construção de uma rede *Future-Proof*. Atualmente a generalidade dos operadores usa a tecnologia *Gigabit Passive Optical Network* GPON definido no organismo de normalização [ITU-T series G.984] mas já se prevê a evolução a curto prazo para XGPON e a médio prazo (2015/2016) para o NG-PON2 (*Next-Generation Passive Optical Network*) sendo imprescindível a reutilização do OSP (*outside plant*). E como será daqui a 10 anos? Haverá necessidade, por exemplo, de redes com rácios de divisão superiores ou inferiores ao que hoje está a ser implementado? Como não é possível ter a certeza, é fundamental planear a rede com alguma flexibilidade, sendo este outro fator imprescindível a considerar desde logo na fase preliminar.

Outro aspeto a ter em conta é a possibilidade de explorar a rede com outro operador. Seja por opção estratégica ou por imposição do regulador nacional esta hipótese tem de ser considerada. Diversos estudos indicam que nas áreas geográficas com maior densidade populacional, o custo por HP diminui favorecendo a decisão de construir uma rede não partilhada (ver Figura 3). Por outro lado, nas zonas rurais torna-se difícil planear um projeto FTTx economicamente viável. Portanto, pode ser vantajoso acordar parcerias em zonas rurais ou na rede interna dos edifícios como forma de reduzir o CAPEX, mas tal pode retirar alguma diferenciação do serviço que é útil do ponto de vista comercial.

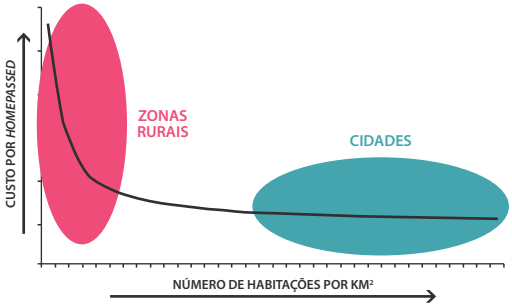


Figura 3. Influência da densidade demográfica no custo/HP

3. FASES DE IMPLEMENTAÇÃO

O processo de implementação de uma rede FTTx é composto por 4 atividades principais: *survey*, projeto, construção e, por fim, cadastro (ver Figura 4). Tratando-se de atividades sequenciais é necessário rigor na sua execução para que os resultados não comprometam as atividades seguintes.

Sendo o *survey* o início de todo o processo, torna-se uma das atividades mais importantes pois um erro nesta fase tem repercussões em todas as seguintes. Quanto mais detalhado e completo tiver sido feito o *survey* menos erros irão surgir no projeto. Dada a importância do *survey*, é conveniente que as pessoas envolvidas sejam especialistas, com capacidade de fazer projeto e conhecimentos de construção. Se assim for, entraves ou dificuldades à construção são detetados antecipadamente e as possíveis soluções a incluir no projeto identificadas desde logo no local. Isso traduz-se numa redução de custos no sentido em que evitará novas deslocações a campo e trabalho adicional decorrente de alterações ao projeto.

Geralmente considera-se a atividade de *survey* como a simples recolha de informações necessárias para o projeto, tais como, infraestruturas existentes, quantificação de unidades de alojamento e características das moradias e edifícios. No entanto, o *survey* deve ser mais do que a recolha de informações quantificáveis. Deve também coletar tendências e particularidades, antecipando necessidades futuras e, desta forma, permitir projetar uma rede *Future-Proof*.

Para otimizar ao máximo o tempo despendido no *survey*, é importante fazer o levantamento da informação num formato compatível com as ferramentas de projeto e cadastro. Para isso, os técnicos de *survey* devem registar a informação coletada de forma georreferenciada em cartografia pré-carregada numa aplicação móvel. Além disso, a importação para as ferramentas de projeto e cadastro deve ser simples e rápida.

Figura 4. Principais atividades na implementação de uma rede FTTx

O projeto tem início quando a informação necessária tiver sido recolhida. Esta é a atividade mais complexa pois recebe os dados coletados no *survey*, incorpora prioridades de atendimento de acordo com a estratégia de venda e marketing da operadora, devolve o custo previsto para essas áreas e define características da rede. Num bom projeto, estas características dependem da densidade demográfica, taxa de penetração prevista, entre outros.

Uma componente elevada do custo de implementar uma infraestrutura de rede FTTx pode ser controlada na atividade de projeto. A título exemplificativo, o investimento inicial de 26% nos equipamentos ativos (ver Figura 2) pode ser reduzido aplicando técnicas de *pay-as-you-grow* tais como o emalhamento de fibras óticas descrito no próximo capítulo. Pode-se portanto dizer que a inteligência de uma rede FTTx reside no projeto e por isso é importante que os projetistas sejam experientes, dominem os materiais e conheçam em detalhe o processo construtivo.

Quando os entregáveis de projeto estiverem concluídos, dá-se início à construção. A chave para uma boa construção à primeira, reside na qualidade do projeto, no conhecimento de quem o executa e no acompanhamento contínuo de obra. É vital ter *feedback* contínuo dos desvios do realizado em relação ao planeado e que o processo seja iterativo. Esta análise diária deve realimentar os processos anteriores tornando-os mais eficientes e previsíveis. Para uma produtividade elevada nesta atividade é importante existir uma boa gestão dos materiais e licenças/autorizações. Após a rede ter sido construída tem de ser validada através de ensaios óticos para aferir alguns dos parâmetros de qualidade da rede: conetividade, atenuações e reflectâncias. Um bom indicador da qualidade de construção é o número de testes de aceitação realizados com sucesso à primeira. Por fim, a construção deve entregar o projeto conforme foi construído para ser cadastrado.

O cadastro é a última atividade, concluindo o processo, não devendo de forma alguma ser descurada. Ao contrário da forma como tipicamente as operadoras encaram as ferramentas de cadastro, estas não devem ser vistas apenas como um custo, dado que têm um papel crucial na diminuição do OPEX. A informação disponibilizada pelas ferramentas de cadastro é preciosa para a operação e manutenção, permitindo um planeamento eficiente dos recursos disponíveis na rede e fundamental para a atuação dos sistemas automáticos de deteção de falhas na redução do tempo de recuperação de serviço. Esta informação permite ainda uma análise comercial que identifique individualmente para cada cliente quais os serviços que lhe podem ser fornecidos.

4. EXEMPLOS DE OTIMIZAÇÃO DO INVESTIMENTO

4.1. EMALHAMENTO DE FIBRAS ÓTICAS

Como referido no capítulo anterior, uma das técnicas para diminuir o CAPEX, sem comprometer a rede em termos de futuro, fiabilidade e robustez é o emalhamento de fibras óticas. Em suma, esta técnica passa por aproveitar a *splitagem* na fibra ótica para distribuir o sinal de um único porto PON por vários pontos de distribuição de rede. Como esta abordagem o investimento ISP (*Inside Plant*) torna-se faseado e dependente da taxa de penetração, contribuindo para a otimização de *hardware* e diminuição de portos ociosos.

4.2. ADEQUADA SELEÇÃO DOS MATERIAIS

Não é percetível numa análise superficial à decomposição do investimento (ver Figura 2) a importância da seleção dos materiais. De facto, apesar dos custos com materiais representar apenas 12% do investimento, a escolha destes influência características de projeto bem como a produtividade na construção. Um bom exemplo é a escolha dos conetores relativamente à sua qualidade. A norma IEC (*International Electrotechnical Commission*) 61300 especifica 4 níveis de qualidade para conetores (ver Tabela 1) dependendo da sua atenuação média em testes normalizados. Evidentemente, os conetores de grade A são mais caros que os de grade D. Contudo, a diferença de atenuações médias é de 0,43 dB permitindo com um único conetor estender a rede por mais 1 Km. Além disso, nos pontos de rede mais distantes da central essa diferença pode significar não ter que reduzir o rácio de *splitagem* o que tem um impacto grande número de portos PON ociosos. Outra vantagem de optar por conetores de boa qualidade é a diminuição das falhas de rede, reduzindo o OPEX.

Attenuation Grade	Attenuation random mated IEC 61300-3-34	
Grade A*	≤ 0.07 dB mean	≤ 0.15 dB max. for > 97% of samples
Grade B	≤ 0.12 dB mean	≤ 0.25 dB max. for > 97% of samples
Grade C	≤ 0.25 dB mean	≤ 0.50 dB max. for > 97% of samples
Grade D	≤ 0.50 dB mean	≤ 1.00 dB max. for > 97% of samples

Tabela 1. Qualidade de conetores segundo norma IEC 61300

Uma outra característica importante dos materiais é serem multifacetados. Se um único modelo permitir várias aplicações diferentes, haverá ganhos de logística, projeto, construção e mesmo manutenção. Um exemplo prático é a possibilidade dos PDOs (Ponto de Distribuição Ótico) permitirem fazer sangria e derivar cabos, o que diminui a quantidade de juntas óticas na rede.

4.3. FERRAMENTAS

No capítulo anterior descreveram-se as principais atividades de um processo de construção de rede FTTx e a forma como estão dependentes umas das outras. Para os processos serem eficazes, eficientes e com menos propensão a erros, é crucial existir uma fonte de informação única para todas essas atividades. Essa ferramenta transversal deve estar disponível não só para o projeto e construção mas também como suporte para as atividades comercial, operação e manutenção.

5. CONCLUSÕES

O investimento necessário para implementar uma rede FTTx depende de múltiplas variáveis, não existindo dois casos iguais. A viabilidade de um projeto FTTx depende fortemente da existência prévia de uma rede de infraestruturas distribuída, do custo de mão-de-obra local e da densidade demográfica das áreas geográficas a atender.

Ao planear uma infraestrutura de rede FTTx as operadoras devem considerar outros fatores além do CAPEX, tais como, o custo de operação e manutenção, a garantia de construção de uma rede *future-proof* e a possibilidade de partilhar a rede com outro operador.

A adoção de uma abordagem *pay-as-you-grow* no projeto, a seleção acertada dos materiais e a escolha de ferramentas transversais a todas as atividades influenciam o sucesso do projeto de construção de uma rede FTTx.

REFERÊNCIAS

[1] "FTTH Business Guide", Second Edition, FTTH Council Europe
[2] "FTTH/B Panorama", IDATE Consulting, FTTH Council Europe Conference, Fev 2013, Londres
[3] "Leading in FTTH – Network Infrastructure", Technology & Innovation Conference, Out 2012, Lisboa
[4] "FTTH Network Economics: Key Parameters Impacting Technology Decisions", 2008, Alcatel-Lucent
[5] "A clear and balanced view on FTTH deployment costs", K. Casier, S. Verbrugge, R. Meersman, D. Colle, M. Pickavet and P. Demeester, 2008, FITCE Congress
[6] http://www.ftthcouncil.eu/

CVS DOS AUTORES

João Figueiredo concluiu, em 1995, *Superior Technician in Electronics* na Universidade de Reading U.K. Ingressou em 2001 na PT Inovação, nos Serviços de Engenharia e desde 2005 que é responsável pela unidade de desenvolvimento e produção de infraestruturas e serviços de FTTx, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede.

Miguel Diz concluiu, em 2008, a Licenciatura em Engenharia Eletrotécnica e de Computadores na Faculdade de Ciências e Tecnologia da Universidade de Coimbra. No mesmo ano ingressou na PT Comunicações através do programa *Trainees* para a área de operação e manutenção da rede de acesso GPON/xDSL. Desde 2011 faz parte da PT Inovação, na direção de Industrialização, Instalação e Suporte de Sistemas de Rede, onde, no inicio de 2012, assumiu funções de *Team Leader* da equipa de serviços de projeto, cadastro e consultoria de redes FTTx.

3

3GPP *3rd Generation Partnership Project*

A

AAA *Authentication, Authorization and Accounting*
AC Ar Condicionado
ACID *Atomic Consistent, Isolated, Durable*
AI *Asset Identification*
AKA *Authentication and Key Agreement*
ANDSF *Access Network Discovery and Selection Function*
ANQP *Access Network Query Protocol*
AP *Acess Point*
API *Application Programming Interface*
APM *Application Performance Monitor*
APN *Access Point Name*
ARF *Asset Reporting Format*
ARP *Address Resolution Protocol*
AMBR *Aggregate Maximum Bit Rate*

B

BSS *Business Support Systems*

C

CCE *Common Configuration Enumeration*
CCSS *Common Configuration Scoring System*
CCTV *Closed Circuit Television*
CDRs *Call Detail Records*
CPE *Common Platform Enumeration*
CS *Circuit Switching*
CVE *Common Vulnerabilities and Exposures*
CVSS *Common Vulnerability Scoring System*

D

DISA *Defense Information Systems Agency*
DSMIP *Dual Stack Mobile IP*

E

E2E *End-to-End*
EAP *Extensible Authentication Protocol-subscriver*
EDCA *Enhanced Distributed Channel Access*
EL *Entreprise Linux*
eNB *evolved Node B*
EPC *Evolved Packet Core*
ePDG *enhanced Packet Data Gateway*
ERP *Enterprise Resourec Planning*
eTOM *enhanced Telecom Operations Map*

F

FIRST *Forum of Incident Response and Security Teams*
FSO *Field Security Operations*

G

GERAN *GSM EDGE Radio Access Network*
GPRS *General Packet Radio Service*
GSMA Global System for Mobile Communications Association

H

HSDPA	*High-Speed Downlink Packet Access*
HSPA	*High-Speed Packet Access*
HS2.0	*Hotspot 2.0*
HSS	*Home Subscriber Server*
HTML	*HyperText Markup Language*
HTTP	*Hypertext Transfer Protocol*

I

IaaS	*Infrastructure-as-a-Service*
I-BCFs	*Interconnection Border Control Function*
ICMP	*Internet Control Message Protocol*
IEEE	*Institute of Electrical and Electronics Engineers*
IMS	*IP Multimedia Subsystem*
IP	*Internet Protocol*
ISMP	*Inter-System Mobility Policy*
ISMS	*Information Security Management System*
ISRP	*Inter-System Routing Policy*
ISO	*International Standards Institute*

J

JSON	*JavaScript Object Notation*

L

LTE	*Long Time Evolution*

M

MAC	*Media Access Control*
MGC	*Media Gateway Controller*
MGCFs	*Media Gateway Controller Function*
MME	*Mobility Management Entity*
MMTEL	*Multimedia Telephony Service*
MO	*Management Objects*
MP	*Manutenções Preventivas*

N

NAT	*Network Address Translation*
NIST	*National Institute of Standards and Technology*
NFC	*Near Field Communication*
NTT	*Nippon Telegraph and Telephone*
NYSE	*New York Stock Exchange*

O

OCIL	*Open Checklist Interactive Language*
OLAP	*Online Analytical Processing*
OLTP	*Online Transaction Processing*
OLT-WZG	*Optical Line Termination - Wireless Zone Gateway*
OMA-DM	*Open Mobile Alliance - Device Management*
OSS	*Operation Support System*
OVAL	*Open Vulnerability and Assessment Language*

P

PaaS	*Platform-as-a-Service*
PAM	*Pluggable Authentication Modules*
PBX	*Private Branch Exchange*
PCC	*Policy Control and Charging*
PCRF	*Policy and Charging Rules Function*
PGW	*PDN Gateway*
PLMN	*Public Land Mobile Network*
PSTN	*Public Switched Telephone Network*

Q

QoS	*Quality of Service*

R

RADIUS	*Authentication Dial In User Service*
RCS	*Rich Communication Suite*
REST	*Representational State Transfer*
RHEL	*Red Hat Entreprise Linux*
RHSA	*Red Hat Security Advisories*
RPM	*Red hat Package Manager*
RTP	*Real Time Protocol*

S

SCAP	*Security Content Automation Protocol*
SCCM	*System Center Configuration Manager*
SDN	*Software-Defined Networking*
SGW	*Serving Gateway*
SLA	*Service-Level Agreement*
SIM	*Subscriber Identity Module*
SLA	*Service-Level Agreement*
SIP	*Session Initiation Protocol*

SNMP *Simple Network Management Protocol*
SOX *Sarbanes-Oxley*
SQL *Structured Query Language*
SSG *SCAP Security Guide*
SSH *Secure SHell*
STIG *Security Technical Implementation Guides*

T

TCO *Total Cost of Ownership*
TCP *Transmission Control Protocol*
TI *Tecnologias de Informação*
TM Forum *Associação da indústria sem fins lucrativos, para provedores de serviços e seus fornecedores nas indústrias de telecomunicações e entretenimento*
TMSAD *Trust Model for Security Automation Data*
TPS *Transações por Segundo*
TTK *Trouble Ticket*
TTLS *Tunneled Transport Layer Security*

U

UDP *User Datagram Protocol*
UE *Equipamento do Utilizador*
UICC *Universal Integrated Circuit Card*
URN *Uniform Resource Name*
USB *Universal Serial Bus*
USGCB *United States Government Configuration Baseline*
USIM *Universal Subscriber Identity Module*
UTRAN *Universal Terrestrial Radio Access Network*

V

VoLTE *Voice over LTE*

W

WAN *Wide Area Network*
WFA *Wi-Fi Alliance*
WFM *Workforce Management*
WiMAX *Worldwide Interoperability for Microwave Access*
WLAN *Wireless Local Area Network*
WPA2 *Wi-Fi Protected Access 2*

X

XCCDF *eXtensible Configuration Checklist Description Format*
XML *Extensible Markup Language*
XMPP *Extensible Messaging and Presence Protocol*

Portugal Telecom Inovação, SA
Rua Eng. José Ferreira Pinto Basto
3810-106 **Aveiro** - Portugal
T.: +351 234 403 200 | **F.**: +351 234 424 723