Universidade Federal de Ouro Preto - UFOP

Instituto de Ciências Exatas e Biológicas - ICEB

Departamento de Computação - DECOM

Caracterização de Padrões de Uso da Rede Sem Fio do DECOM
Relatório de Atividades Desenvolvidas em Monografia 1.

Aluno: Paulo Henrique Campos de Freitas
Matrícula: 07.1.4050

Orientador: Daniel Fernandes Macedo

Ouro Preto
3 de dezembro de 2010

Universidade Federal de Ouro Preto - UFOP

Instituto de Ciências Exatas e Biológicas - ICEB

Departamento de Computação - DECOM

Caracterização de Padrões de Uso da Rede Sem Fio do DECOM
Relatório de Atividades Desenvolvidas em Monografia 1.

Relatório de atividades desenvolvidas apresentado ao curso de Bacharelado em Ciência da Computação, Universidade Federal de Ouro Preto, como requisito parcial para a conclusão da disciplina Monografia I (BCC390).

Aluno: Paulo Henrique Campos de Freitas
Matrícula: 07.1.4050

Orientador: Daniel Fernandes Macedo

Ouro Preto
3 de dezembro de 2010

Resumo

As redes sem fio são cada vez mais comuns hoje em dia. Uma rede mal projetada pode ser lenta e não satisfazer seus usuários à medida em que eles aumentam. Para melhorar seu desempenho, uma rede deve ser medida a fim de identificar pontos críticos em sua implementação. O trabalho a seguir visa medir os pacotes trafegados pela rede sem fio do Departamento de Ciência da Computação da Universidade Federal de Ouro Preto. A partir da medição será possível identificar pontos de melhorias para a rede, assim como avaliar os serviços oferecidos por ela.

Palavras-chave: Rede sem fio. Coleta de dados. Mikrotik. Anonimização.

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

Nos últimos anos, o número de alunos do Departamento de Computação (DECOM) da Universidade Federal de Ouro Preto que possuem notebooks cresceu consideravelmente. Consequentemente, o uso destes nos laboratórios da universidade propiciou um significativo aumento de perdas de mouse, cabos de redes, e outros periféricos que causaram prejuízos à universidade. Para minimizar essas perdas, foi proibido pelo DECOM o uso de notebooks nos laboratórios.

De forma a não prejudicar os alunos, foi implantado pelo DECOM, sob a coordenação do professor Ricardo Rabelo, uma rede sem fio. Esta é restrita aos alunos e professores, os quais podem acessá-la através de pontos de acesso espalhados pelo departamento e no Centro Acadêmico de Ciência da Computação (CACIC), mediante autenticação.

A topologia da rede sem fio consiste em uma máquina servidora e vários pontos de acesso. Existem ainda dois roteadores Mikrotik [3], porém ainda não estão em uso e futuramente serão instalados.

Os pontos de acesso são basicamente compostos por computadores simples, sem grandes quantidades de processamento e memória, e com discos rígidos razoáveis, de modo a possibilitar a reciclagem de máquinas consideradas ultrapassadas. Estas máquinas possuem um sistema operacional Mikrotik RouterOS, que oferece algumas ferramentas para monitorar a rede, como geração de arquivos de *log*, e um *sniffer* para rastrear os pacotes trafegados na rede, além de outras que possibilitam transformá-las em roteadores. A máquina servidora possui configurações melhores, uma vez que máquinas similares aos pontos de acesso queimaram quando utilizadas para esta finalidade.

A instalação da rede é uma experiência recente, e ainda não foi realizada nenhuma medição de tráfego dessa.

Medir o tráfego da rede consiste em coletar e identificar os dados que trafegam por essa rede além de analisá-los e caracterizá-los de forma a identificar padrões. Os dados que não se encaixem nos possíveis padrões identificados também devem ser separados e analisados. Os dados podem ser separados entre conexões HTTP, serviços de e-mail, conexões peer-to-peer, entre outras possíveis categorias que podem ser identificadas durante a análise dos resultados. A caracterização da rede será feita a partir desses dados, e se faz necessária para que seja possível identificar pontos de melhoria dessa e propor soluções para tais. A análise dos dados propiciará também um balanço a respeito do uso da rede e de seus serviços.

Todos os dados coletados serão anonimizados de forma a ocultar o usuário e preservar sua identidade, afinal não é objetivo do trabalho apontar quais usuários estão erroneamente utilizando a rede.

As seções seguintes apresentarão a proposta para o trabalho de Monografia

CARACTERIZAÇÃO DE PADRÕES DE USO DA REDE SEM FIO DO DECOM. Seu objetivo é montar um ponto de acesso de teste, medir e analisar o tráfego da rede do DECOM, assim como os serviços disponibilizados por ela como o Moodle, o servidor Web, entre outros. Desta forma será possível detectar falhas na rede e indicar soluções para melhorar a velocidade e disponibilidade da rede e seus serviços para todos os seus usuários. A Seção 2 explica as justificativas para a escolha do trabalho. Em seguida a Seção 3 apresenta os objetivos gerais e específicos do trabalho. Na Seção 4 se encontra a metodologia a ser utilizada para a execução deste e por fim, a Seção 5 apresenta o cronograma para seu desenvolvimento.

2 Justificativa

A motivação para o desenvolvimento deste trabalho é conhecer melhor a rede sem fio, de modo que seja possível indicar melhorias para a mesma e para seus serviços.

2.1 Medindo a rede

Antes de poder sugerir qualquer adaptação para a rede sem fio, é necessário conhecer como ela funciona, o que é transmitido através dela. De modo geral, é preciso analisar o ponto de acesso para identificar possíveis melhorias.

A forma adotada para conhecer o ponto de acesso será a medição dos pacotes trafegados por ele. Depois de feita a coleta, os dados serão separados e analisados, possibilitando identificar pontos de melhoria na rede e até possíveis ataques à rede.

2.2 Serviços da rede

São vários os serviços disponibilizados pela rede sem fio. Medir esses serviços é a maneira adequada de identificar melhorias viáveis. Entre as melhorias, pode-se analisar a segurança destes: A comunicação do usuário com o Moodle é segura? Os e-mails são protegidos contra interceptadores? O servidor WEB é capaz de defender os usuários contra vírus? Além da segurança, pode-se medir a eficiência de tais serviços, como a taxa de utilização, velocidade de resposta, entre outras características.

3 Objetivos

3.1 Objetivo geral

- Estudar os dados trafegados pela rede sem fio é fundamental para identificar seus pontos críticos. Uma vez identificados, será possível sugerir melhorias. Este trabalho irá estudar a rede sem fio e os serviços disponibilizados por ela e propor melhorias.

3.2 Objetivos específicos

- Redes de computadores são cada vez mais frequentes no dia a dia, principalmente agora com a expansão da internet. Este trabalho irá revisar conceitos relacionados à construção e administração de redes de computadores.

- Surgem a cada dia diversas ferramentas para gerenciar redes de computadores. Neste trabalho será possível conhecer melhor algumas destas que o Mikrotik RouterOS oferece.

- O Mikrotik RouterOS é um sistema operacional desenvolvido pela empresa Mikrotik [3] que permite transformar um simples computador em um poderoso roteador. Ele será apresentado com suas ferramentas ao longo desse trabalho.

- Para analisar o tráfego da rede é necessário medir todas as informações trafegadas pelo ponto de acesso desta. Para isso será montado um ponto de acesso de teste na rede sem fio e desta forma obter os arquivos de *log*.

- Para proteger a privacidade dos usuários da rede os dados coletados devem ser anonimizados. A técnica que será adotada é a anonimização com preservação de prefixo. Ela será detalhada e apresentada ao longo do trabalho.

- As informações úteis para a análise estarão registradas nos arquivos de *log*. Para auxiliar a análise será desenvolvido um protótipo responsável por processar os arquivos e extrair deles as informações desejadas

4 Metodologia

Nesta seção será apresentada primeiramente uma revisão da bibliografia utilizada para o desenvolvimento do trabalho, definindo os algoritmos e ferramentas que serão utilizados durante a disciplina de Monografia I (BCC 390). Em seguida são detalhadas as etapas do desenvolvimento para a disciplina de Monografia II (BCC 391).

4.1 Revisão Bibliográfica

A revisão bibliográfica apresentará uma revisão de conceitos de redes de computadores encontrados em [4], uma vez que a disciplina de Redes de Computadores será cursada pelo autor em paralelo ao desenvolvimento do trabalho no próximo semestre.

A troca de dados entre computadores de uma rede é feita através de mensagens. Essas estão encapsuladas dentro de pacotes, que nada mais são do que uma sequência de *bytes* [4]. Uma mensagem pode ser subdividida em vários pacotes, e são inúmeros os pacotes trocados em uma comunicação. Para medir uma rede, é necessário fazer uma coleta dos dados trafegados por ela. A coleta pode ser feita de diversas maneiras, entre elas a por interceptação. Esse modelo apenas adiciona uma funcionalidade ao ponto de acesso, que é a de registrar os pacotes que passam por ele.

Existem ferramentas desenvolvidas para executar a coleta dos pacotes no ponto de acesso. Uma delas é o Ethereal desenvolvida pela Ethereal [1]. Ela é uma ferramenta de código livre desenvolvida para executar nos principais sistemas operacionais existentes no mercado, entre eles Unix, Linux e Windows. Existe também o Wireshark [6], que é a evolução do Ethereal.

A empresa Mikrotik desenvolveu um sistema operacional chamado Mikrotik RouterOS. Ele é baseado em linux e possui inúmeras ferramentas para auxiliar na construção e administração da rede. O site da Mikrotik disponibiliza uma seção wiki [5] para apresentar o Mikrotik RouterOS e todas as ferramentas oferecidas por ele. Nela também é possível encontrar todas as especificações para o WinBox.

Para proteger a privacidade do usuário deve-se anonimizar o pacote coletado. Essa anonimização pode ser feita nos campos IPs de destino e origem. Para mascarar um IP existe a técnica de anonimização com preservação de prefixo descrita em [7]. Nesta técnica, se temos dois endereços IP com *bits* iniciais iguais, após anonimizados, eles terão a mesma quantidade de *bits* ligados iguais.

4.2 Desenvolvimento

O primeiro passo para o desenvolvimento do trabalho será montar um ponto de acesso de teste e disponibilizá-lo para os alunos do DECOM, ou caso seja necessário devido a pouco acesso, para os demais alunos do ICEB.

O ponto de acesso será constituído de um computador com quantidade razoável de memória RAM e espaço de armazenamento. Seu papel será transmitir os pacotes da

máquina servidora para os possíveis clientes que estejam conectados. Não é uma tarefa que demanda muito poder de processamento, porém caso queira processar os pacotes *online*, é necessário um pouco mais de recurso computacional. Para esse trabalho iremos gerar um arquivo de *log* no ponto de acesso e processar esse arquivo em outra máquina, de modo a exigir do ponto de acesso apenas uma quantidade significativa de memória de armazenamento.

Inicialmente a coleta ocorerá apenas neste ponto de acesso para que seja avaliado se ela irá prejudicar o desempenho da máquina. Uma vez realizados os testes de desempenho, pode-se expandir a coleta para todos os pontos da rede, gerando uma quantidade de dados maior. Desta forma podemos medir todos os usuário da rede, evitando o risco de coletar dados de apenas um grupo específico de usuários.
Cada ponto de acesso executa o sistema operacional Mikrotik RouterOS. O sistema é desenvolvido para gerenciar as Mikrotik RouterBoards, porém se instalado em qualquer computador, possibilita transformá-lo em um roteador. A interface com o usuário é através de linha de comando, porém existe ainda uma ferramenta chamada WinBox que permite acessar todas as funcionalidades oferecidas pelo Mikrotik por meio de interfaces gráficas. A vantagem em se usar esta ferramenta é que ela facilita monitorar os dados que trafegam pela rede, e não precisa ser executada na própria máquina, podendo ser acessada em qualquer computador que esteja conectado à rede. Por motivos de segurança toda a comunicação entre o Mikrotik e o WinBox é criptografada.

A coleta dos dados será feita por meio da ferramenta Traffic Flow [5], disponibilizada pelo Mikrotik Router OS. Essa ferramenta é capaz de coletar todos os pacotes que trafegam através do ponto de acesso. Traffic Flow é baseado na ferramenta Cisco NetFlow e proporciona todas as utilidades desta. A partir da coleta dos dados, é necessário anonimizar os dados, de forma a proteger a privacidade dos usuários da rede.

Anonimizar um dado consiste em mascarar sua identidade, de forma a tornar impossível a identificação de seu autor. Existem leis que se referem ao fato de preservar a privacidade e segurança do usuário, portanto é necessário escolher cautelosamente qual ferramenta utilizar e qual política adotar. A ferramenta de anonimização escolhida deve garantir que os dados do mesmo usuário sejam agrupados e analisados, porém não pode permitir nenhuma inferência entre os dados anonimizados com dados reais.

No caso deste trabalho, iremos usar a anonimização de endereço IP. O endereço IP permite identificar a origem de um pacote e seu fluxo na rede, sendo de grande valia na análise a ser feita. A técnica de anonimização utilizada será a preservação de prefixos, pois permite agrupar os dados pertencentes a uma mesma sub-rede porém preservando a privacidade do usuário.

Após a etapa de anonimização dos dados, serão feitos os processamentos e análise dos dados. Um protótipo será gerado para analisar os arquivos de *log* gerado pela ferramenta Traffic Flow e agrupar os pacotes similares. O protótipo irá gerar as estatísticas de todo o tráfego, levando em conta qual serviço está sendo utilizado, possibilitando deduzir quais desses serviços podem ser otimizados.

5 Desenvolvimento

Nesta seção, serão descritas as atividades desenvolvidas ao longo do período para a disciplina BCC 390 Monografia I.

No início do desenvolvimento desse trabalho, o autor não possuía conhecimentos suficientes sobre redes de computadores pois ainda não cursou a disciplina de Redes. A disciplina será cursada por ele durante o próximo período, paralelamente ao desenvolvimento da monografia. Para iniciar o trabalho, foi desenvolvido ao longo do período um profundo estudo a respeito de redes de computadores. Com base nesse estudo, foi possível identificar os principais protocolos de comunicação para redes e com isso identificar como cada um pode contribuir para o trabalho. Como o ponto de acesso a ser utilizado possui o Mikrotik RouterOS como seu sistema operacional, foi necessário um estudo do sistema com a finalidade de identificar o que ele tem a oferecer para monitorar o tráfego da rede. A seguir será apresentado os resultados obtidos.

5.1 Mikrotik RouterOS

O sistema operacional Mikrotik RouterOS foi desenvolvido para permitir que um computador possa disponibilizar serviços de um roteador. Ele é capaz de oferecer outros serviços como *firewall*, controle de banda, além de servir como ponto de acesso para redes sem fio. As ferramentas de *log* do Mikrotik registram informações a respeito dos usuários durante o momento em que esses estão conectados, podendo essas serem agrupadas por pacotes, protocolos, ou por conexões. Como exemplo, os *logs* capturados por pacotes incluem o endereço de origem e destino, o protocolo utilizado e o tamanho dos pacotes. Caso seja necessário fazer uma coleta mais específica onde seja possível ter acesso a todo o conteúdo do pacote, deve ser utilizado uma ferramenta para auxiliar. Nesse caso, o Wireshark é a melhor opção.

5.2 Anonimização dos dados

A anonimização de endereço IP em uma coleta de tráfego de rede é indispensável para garantir a privacidade do usuário da rede. Quando é desejável saber os endereços pertencente a uma mesma sub-rede, a anonimização por preservação de prefixos se faz necessária, pois ela permite que dois IPs que compartilham k bits iguais, quando anonimizados também compartilham os mesmos k bits. Esse método pode ser vulnerável a ataques, e para aumentar a segurança, [7] propôs a anonimização de endereço IP com preservação de prefixo baseado em criptografia. A garantia de segurança do método esta ligada ao fato do uso das funções de *hash* em sua execução. A função de *hash* é responsável por garantir que um determinado número se mantenha igual toda vez que ela seja aplicada a este número. O uso de chaves para algoritmos baseados em *hash* permite que os arquivos de *log* sejam separados e processados separadamente, garantindo os resultados iguais, caso sejam utilizadas cópias idênticas da chave utilizada.

5.3 Protocolo TCP/IP

Internet Protocol ou IP é um protocolo da camada de rede para endereçamento e roteamento de dados. Os pacotes IP possuem um cabeçalho onde são definidos sua versão, tamanho do cabeçalho e tamanho total do pacote (cabeçalho e dados),

protocolo de transporte, endereço IP de origem e destino, além de algumas outras opções como a lista de roteadores por onde um determinado pacote deve passar até chegar a seu destino. Alguns campos adicionais podem ser necessários quando o pacote chegar à uma determinada rede que precise fragmentá-lo para transportá-lo.

Atualmente a versão mais usada do protocolo IP é o IPv4, composto por quatro números separados por ponto e somando um total de 32 bits. Teoricamente, cada estação conectada à Internet deve possuir um endereço único e confiável. No entanto com o avanço da Internet é cada vez mais evidente a necessidade de uma nova versão, afinal é grande a quantidade de novos usuários. A adoção do IPv6 está acontecendo aos poucos e é a atual proposta para a solução deste problema.

O endereço IP pode ser divido permitindo identificar o endereço da rede e do cliente. Essa divisão permite otimizar a comunicação entre estações integrantes de uma mesma sub-rede, evitando a necessidade de enviar os pacotes para o roteador e transferindo-os pelo próprio barramento de redes. A indentificação de clientes pertencentes à uma sub-rede em comum ocorre através de uma multiplicação binária entre o endereço do cliente e uma máscara de rede. Caso a multiplicação possua resultados iguais para os dois endereços, ambos pertencem a mesma sub-rede. O endereço IP ainda é dividido em classes diferentes de acordo com o número de sub-redes e *hosts* que cada uma possue.

TCP ou *Transmission Control Protocol* é um protocolo para a camada de transporte confiável e orientado a conexões. Confiável pois, ele garante a entrega dos pacotes IP ao destinatário, além de ser o responsável por separar e reagrupar os dados trocados entre as aplicações. A comunicação do TCP é feita ponto a ponto, ou seja, é necessário que exista uma estação de origem e uma de destino. A conexão TCP pode ser identificada pelo par endereço e porta, onde as portas são números inteiros de 16 bits que identificam a qual aplicação o pacote TCP deve ser entregue. Existem 65.536 portas TCP, o que em teoria nos permite ter a mesma quantidade de aplicações se comunicando simultaneamente. As portas numeradas de 1 a 1024 são registradas pela *Internet Assigned Numbers Authority* (IANA) [2] e reservadas para protocolos da camada de aplicação como por exemplo a porta 80 para HTTP, porta 20 e 21 para FTP e a porta 25 para SMTP.

A identificação da porta utilizada em uma conexão permite descobrir qual aplicação está comunicando em determinado momento, porém há casos em que apenas o número da porta não é válido, como por exemplo as redes *peer-to-peer* (P2P) atuais. Aplicações P2P conseguem se conectar através da porta 80; técnica eficiente para mascarar a conexão fazendo se passar por um acesso à um servidor WEB. Casos como esse exigem que sejam analisados mais dados dos pacotes comutados.

6 Trabalhos Futuros

Os trabalhos futuros serão executados no próximo período juntamente com a escrita da monografia. Eles podem ser divididos em três etapas listadas a seguir:

6.1 Aquisição dos dados

Para que se tenha uma boa análise é necessário possuir uma quantidade significativa de dados. Para coletar os dados será montado um ponto de acesso utilizando o Mikrotik RouterOS que ficará disponível aos alunos, esses serão informados sobre a finalidade deste ponto de acesso e a importância de medir a rede para indicar melhorias. Devido a alta velocidade na aquisição dos pacotes, os arquivos de *log* podem ser divididos, facilitando a etapa de análise dos dados. Ainda no fim deste semestre será montado o ponto de acesso para que o autor possa se familiarizar com o Mikrotik, uma vez que esse só possui conhecimentos teóricos sobre o sistema. Com o ponto montado e pronto para iniciar a coleta, espera-se que nas primeiras semanas de aula do próximo período os dados para o início da analise se encotrem à disposição.

6.2 Análise dos dados

Antes de ser feita a análise dos dados, será necessário processar os arquivos de *log* adquiridos de forma a anonimizar os endereços IP de origem e destino dos pacotes. Ainda deverá ser definido se mais alguma informação deve ser descartada, pois não se pode descartar a possibilidade de interceptar dados pessoais como senhas, documentos e outras informações que possam prejudicar o usuário e afetar sua privacidade. Uma vez que os arquivos de *log* estejam prontos para serem analisados, será feito um levantamento dos dados procurando identificar o comportamento dos usuários da rede. A técnica de anonimização escolhida permite identificar pacotes diferentes pertencentes ao mesmo usuário sem, portanto, revelar sua identidade, além de garantir que pacotes pertencentes à uma mesma sub-rede possam ser agrupados mesmo depois de anonimizados. Dessa forma é possível medir o tráfego de um usuário, sem ter que identificá-lo, e até mesmo o tráfego de uma sub-rede.

6.3 Apresentação dos resultados

Depois que todos os arquivos de *log* forem analisados deverão ser confeccionados gráficos e tabelas de forma a ilustrar os resultados obtidos. Com esses resultados prontos, será possível identificar um padrão de uso da rede do DECOM e o comportamento dos usuário da rede. É desejável descobrir gargalos que causem lentidão à rede e sugerir métodos para contorná-los. A segurança da rede é mais um ponto a ser analisado, afinal é possível identificar alguns tipos de ataques através dos pacotes trafegados. Para todos os problemas identificados pela análise serão feitas sugestões de melhorias ou formas de tentar contorná-los. Todos os resultados serão devidamente documentados durante a redação da monografia.

7 Cronograma de atividades

Na Tabela 1, será apresentado o cronograma para o desenvolvimento da disciplina Monografia 2.

Atividades	**Fev**	**Mar**	**Abr**	**Mai**	**Jun**
Coleta do dados	X	X			
Análise dos dados	X	X	X		
Apresentação dos resultados			X	X	
Redigir a Monografia			X	X	X
Apresentação do Trabalho					X

Tabela 1: Cronograma de Atividades.

Referências

[1] Site oficial do ethereal. `http://www.ethereal.com`. Visitado em 20 de Outubro de 2010.

[2] Site oficial da internet assigned numbers authority. `http://www.iana.org`. Visitado em 27 de Outubro de 2010.

[3] Site do mikrotik. `http://www.mikrotik.com`. Visitado em 16 de Outubro de 2010.

[4] Andrew S. Tanenbaum. *Redes de Computadores*. Campus, 4 edition, 2003.

[5] Site wiki do mikrotik. `http://wiki.mikrotik.com/`. Visitado em 16 de Outubro de 2010.

[6] Site do wireshark. `http://www.wireshark.org`. Visitado em 20 de Outubro de 2010.

[7] Jun Xu, Jinliang Fan, Mostafa Ammar, and Sue B. Moon. On the design and performance of prefix-preserving ip traffic trace anonymization. In *ACM SIGCOMM Internet Measurement Workshop*, pages 263–266, San Francisco, USA, 2001.