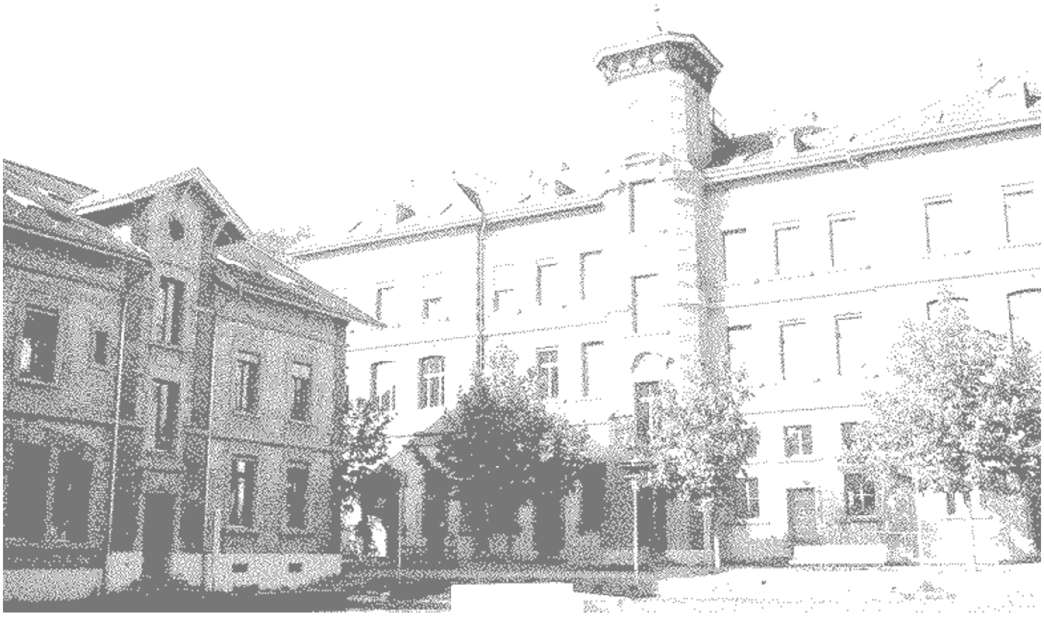
Doctoral Workshop on Distributed Systems

H. Mercier, T. Braun, P. Felber, P. Kropf, P. Kuonen, E. Rivière (eds.)

Technical Report IAM-12-003, September 10, 2012

Institut für Informatik und angewandte Mathematik, www.iam.unibe.ch



Doctoral Workshop on Distributed Systems

Hugues Mercier, Torsten Braun, Pascal Felber, Peter Kropf, Pierre Kuonen, Etienne Rivière

Technical Report IAM-12-003, September 10, 2012

CR Categories and Subject Descriptors:
C.2.1 [Computer-Communication Networks]: Network Architecture and Design; C.2.2 [Computer-Communication Networks]: Network Protocols; C.2.3 [Computer-Communication Networks]: Network Operations; C.2.4 [Computer-Communication Networks]: Distributed Systems

General Terms:
Design, Management, Measurement, Performance, Reliability, Security

Additional Key Words:
Mobile communication networks, wireless mesh networks, wireless sensor networks, cloud computing, localization, time synchronization, fault tolerance, Internet traffic anomalies, software transactional memories, privacy, DNA sequencing

Institut für Informatik und angewandte Mathematik, Universität Bern

Abstract

The Doctoral Workshop on Distributed Systems has been held at La Vue-des-Alpes, Switzerland, from July 4-6, 2012. Ph.D. students from the Universities of Neuchâtel and Bern as well as the University of Applied Sciences of Fribourg presented their current research work and discussed recent research results. This technical report includes the extended abstracts of each talk given during the workshop.

VUE-DES-ALPES 2012

4 – 6 JULY

DOCTORAL WORKSHOP ON DISTRIBUTED SYSTEMS

UNIVERSITÄT BERN

UNIVERSITÉ DE NEUCHÂTEL

Workshop Program

Wednesday, June 4

14:00 Reliable Communication in Energy Efficient WSNs
Markus Anwander

14:25 Multimedia Intrusion Detection in Multi-tier Wireless Multimedia Sensor Networks
Denis do Rosàrio

14:50 Accounting in Wireless Mesh Networks
Alexey Monakhov

15:15 Coffee Break

15:45 A Unified Technique to QoS Guaranteed Scheduling in LTE-Advanced Networks – Emphasis on System Throughput and User Fairness
Ioan S. Comsa

16:10 Software Defined Radio and USRP Synchronization with GPS
Zan Li

16:35 Routing, Coverage and Connectivity Issues in UAV ad-hoc Networks
Zhongliang Zhao

Thursday, June 5

9:00 Dynamic SLA Management for Infrastructure-as-a-Service Cloud Environments
Alexandru-Florian Antonescu

9:25 StreamHub: Scalable Content-Based Routing for Event Streams
Raphaël Barazzutti

9:50 Thrifty Privacy: Efficient Support for Privacy-Preserving Publish/Subscribe
Emanuel Onica

10:15 Coffee Break

10:45 Integrating IoT and Enterprise IT: On the Need of Semantic Descriptions for Interoperability
Matthias Thoma

11:10 Decentralized, Scalable and Trusted Storage & Indexing
José Valerio

Friday, June 6

9:00 Indoor Localization and Dynamic Environment Detection
Islam Alyafawi

9:25 Fault Tolerance Based on POP-C++
Jianping Chen

9:50 Analysis of Anomalies in the Internet Traffic
Veronica del Carmen Estrada Galinanes

10:15 Coffee Break

10:45 Supporting Time-Based QoS Requirements in Software Transactional Memory
Walther Maldonado

11:10 Distributed DNA Alignment, a Stream-Based Approach
Beat Wolf

Presentation Abstracts

Indoor Localization and Dynamic Environment Detection

Islam Alyafawi, Universität Bern
alyafawi@iam.unibe.ch

Abstract

Indoor localization will be an inseparable feature of mobile services/applications in future wireless networks. The current ubiquitous availability of indoor localization systems is still obstructed by technological challenges and privacy issues. We propose an innovative approach towards the concept of indoor positioning with the main goal to develop a system that is self-learning and able to adapt to various radio propagation environments. The approach combines estimation of propagation conditions, subsequent appropriate channel modeling and feedback optimization to the used positioning algorithm. Main advantages of the proposal are decreased system set-up effort, automatic re-calibration and increased precision.

Keywords: propagation models; localization techniques; indoor applications.

1 Introduction

Wireless applications have been rapidly developed and its importance is continuously increasing. Indoor environments present opportunities for a rich set of location-aware applications such as navigation tools for humans and robots, resource discovery, location-aware sensor networking etc. Typical indoor applications require better accuracy than what outdoor location systems provide, as in (museums, shopping malls, hospitals). The global positioning system (GPS) technology has solved the problem of outdoor localization, but an effective and affordable localization solution for indoor environments is still challenging to find [1]. A localization solution combines input parameters from a specific technology with a positioning algorithm that processes these parameters to derive location coordinates of a target object.

Technologies. Several competing technologies are discussed for localization purposes. *Bluetooth and WiFi* are attractive due to their wide spread adoption and low deployment cost. However, location is derived from radio signal strength and therefore radio interference introduces position inaccuracies. *Radio-Frequency IDentification (RFID)* technology is easy to deploy with cheap devices but it requires high installation density [3]. *Ultra-wide band (UWB)* can provide centimeter-based accuracy and robustness to interference at the cost of expensive hardware needed for accurate time synchronization [6]. Cellular networks are also well studied for the purpose of support localization, especially GSM technology, due to GSM's ubiquitous coverage

Positioning techniques. The number of proposed localization techniques is large as well. The most often cited techniques are angulation, lateration and fingerprinting [1, 6]. The choice of techniques is determined by the environment characteristics, cost and complexity of the localization equipment. In an *angulation/lateration* technique the target coordinates are derivatives of different angle/distance measurements to known reference points [6]. Distances can be obtained by various methods such as time of arrival (ToA), time difference of arrival (TDoA), or received signal strength (RSS). Lateration accuracy depends on the input parameters and their vulnerability to indoor multipath propagation. A *fingerprinting* technique compares real-time measurements to an off-line constructed database [2]. Since the database is constructed for a specific indoor environment, fingerprinting is time and effort consuming for dynamic environments [7, 5]. An improvement to the technique explored by [7] is to alternate between several pre-generated databases depending on sensed changes in the radio channel.

Adaptive localization system (ALS). We take an alternative approach towards indoor positioning, namely a flexible system that can dynamically detect changes in the radio environment and adapt to them. ALS focuses on the use of radio frequency technologies such as WiFi and Bluetooth with an extension towards GSM/UMTS technologies. The goal of ALS is easy deployment, minimum re-calibration and high localization accuracy.

2 Adaptive Localization System

The proposed technique consists of two components. The first component - an adaptive propagation model - is responsible for monitoring the indoor propagation conditions and adapting the parameters of the propagation model when necessary. The second component - a self-learning localization algorithm - uses feedback from the (adapting) propagation model to re-calibrate the localization mechanism.

2.1 Adaptive Propagation Model

Radio-based systems for indoor localization that rely on RSS strongly depend on appropriate modeling of the propagation channel. An indoor radio channel is highly affected by multipath components and shadowing due to complex indoor layout, e.g., multitude of obstacles on the signal path. Therefore, the accuracy of the whole system depends on how well the different signal components (reflected, scattered and refracted) are modeled.

Indoor radio channels have been popularly modeled by a log-normal shadowing model [4]. Most previous works on indoor propagation models apply fixed propagation parameters such as shadowing variance σ^2_{dB} and path loss exponent α. A disadvantage of such approach is its inability to scale to different indoor structures. Each indoor structure has its own specific characteristics. Moreover, changes in the structure (e.g., moving persons) causes changes in the propagation parameters.

On the other hand, systems relying on TDoA are more robust to indoor layout. However, the time resolutions provided by many sensor nodes are limited in resolution, and do not achieve the required accuracy for indoor applications.

Our target is to develop a smart algorithm to derive propagation parameters without any pre-knowledge on the channel. We rely on periodically listening to the channel between reference points with known location, to gain knowledge on the propagation conditions and changes therein. A similar approach but only applied to fingerprinting is proposed in [5]. The WiFi radio channel is periodically sensed to detect changes in the propagation conditions and to correct the fingerprinting database. The authors show promising results that can solve the problem of re-calibration. Our goal is to explore even further the potential of ALS to combine different localization schemes and to improve the localization accuracy.

2.2 Self-learning Localization Algorithm

We plan to investigate the benefits of the adaptive model for two positioning techniques. Our target is to use the proposed adaptive propagation model to minimize distance errors by updating the model parameters either periodically or before a positioning decision is made. As a result of the update a more up-to-date RSS estimation can be made.

As an additional improvement we envision to combine the adaptive propagation model with opportunistic techniques for signal derivation. One such technique is the careful selection of which reference signals to use in the localization decision (or to acquire the new propagation parameters for that matter) and the weighting of each contribution when deriving an object's coordinates. Another technique is to make use of channel diversity by collecting readings on several channels and taking the combined feedback in the RSS evaluation. The latter approach is inspired by [8].

Additionally, advantages are also foreseen for localization methods using ANN. Training the network to identify a correct position can be very time consuming. We believe that our proposed model can be improved by providing more accurate data as input and thus decrease learning times.

References

[1] K. Al Nuaimi and H. Kamel. A survey of indoor positioning systems and algorithms. In *International Conference on Innovations in Information Technology (IIT)*, pages 185 –190, 2011.

[2] P. Bahl and V.N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775 –784 vol.2, 2000.

[3] G.Y. Jin et al. An indoor localization mechanism using active rfid tag. *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing.*, 2006.

[4] J. Xu et al. Distance measurement model based on rssi in wsn. *Wireless Sensor Network*, 2:606–611, 2010.

[5] L. Moraes et al. Calibration-free wlan location system based on dynamic mapping of signal strength. *MobiWac*, pages 92–99, 2006.

[6] M. Mohamed et al. Investigation of high-accuracy indoor 3-d positioning using uwb technology. *IEEE Transactions on Microwave Theory and Techniques*, 56, 2008.

[7] V. Bahl et al. Enhancements to the RADAR user location and tracking system. Microsoft Research.

[8] P. Pettinato, N. Wirström, J. Eriksson, and Th. Voigt. Multi-channel two-way time of flight sensor network ranging. to appear in EWSN 2012.

Dynamic SLA Management for Infrastructure-as-a-Service Cloud Environments

Alexandru-Florian Antonescu, Universität Bern, SAP Research
antonescu@iam.unibe.ch

Cost-efficient operation while satisfying performance and availability guarantees in Service Level Agreements (SLAs) is a challenge for Cloud Computing, as these are potentially conflicting objectives. This work describes a framework for SLA management that addresses this challenge by featuring a forecasting model for determining the best virtual machine-to-host allocation given the need to minimize SLA violations, energy consumption and resource usage.

A Service Level Agreement (SLA) is a contract between a consumer and a provider of a service regarding its usage and quality. From the perspective of the provider, SLAs describe the resource kinds, QoS terms and pricing policy, while from the consumer perspective, SLAs describe the amount of resources to be used and QoE terms. Providers will seek ways of ensuring optimal SLA satisfaction, within the constraints of their resource usage, energy consumption and budget constraints, in order to maintain profitable businesses. Having insight or prior knowledge of application workloads and data volumes that will be running on their infrastructure is hence beneficial towards tuning their infrastructure for optimizing both of these objectives. Using an effective, automated SLA management architecture should be considered important in this context. When prioritizing operational tasks and resources based on their commitments, capacities and capabilities, providers then use the information in an SLA. Providers aim to satisfy the Service Level Objectives of consumers without disrupting their internal operation goals including minimization of operational costs, power consumption, and legal issues. The demand for rapid delivery of highly available, high performance, networked applications and infrastructure services makes automation support for SLA Management more critical.

Existing and emerging application and infrastructure services no longer represent single, isolated, remote servers or virtual machines (VMs). Disruptions in one resource can have consequences for the overall composite service, such that state information needs to be controllably shared. There are then dependencies across operational layers (application, machine and network) and domains with autonomous administrators, lifecycles and management controllers, which can lead to a break in the autonomy property. The application topologies continue to become more complex and heterogeneous as service-oriented design is used. For these reasons the inter-dependencies between application deployment, usage and network capabilities need to be considered during service planning and provisioning, making the coordination of SLA management more complex.

The SLA management problem can be summarized as a four dimensional problem, such that a solution must take these four dimensions into account: - Multiple layers of technology, such as physical resource access, virtualization, infrastructure services and applications, which contribute to the service execution. - Multiple resource types (compute, storage, memory and network) are required and aggregated for the purpose of delivering the service. - There are multiple entity organizations (providers, brokers, consumers) acting in a value chain of providers and consumers of resources and services. - There are multiple objectives and metrics to be combined within SLAs, which may be conflicting with each other and have different priorities and significance across layers (applications, infrastructure services, physical resource layer). SLA management includes multiple functions starting with modeling the computing environment's resources and capabilities, using SLA terms, followed by negotiation of the customer QoE terms and resources. Next, the resource allocation and planning are orchestrated onto the available physical resources followed by monitoring the runtime behavior of the virtual infrastructures and reacting to SLA violations. The cycle is then completed with auditing the finished SLA contracts.

As such, the problem domain can be described as dealing with on-demand services that use shared infrastructure with virtualization capabilities (e.g. VMs running on physical hosts). Virtual machines and application workloads can be scaled dynamically determining that the infrastructure needs to be responsive & dynamically provisioned, also with regards to network management. The service level objectives (e.g. high availability, high performance) often conflict with the provider's operational objectives (e.g. energy efficiency, high profits). Here we focus primarily on conflicts of service level objectives with energy efficiency. The overhead of virtualization management necessary for maintaining service level objectives (e.g. VM migration) needs to be considered when planning the infrastructure utilization.

The current work focuses on producing a SLA Management system capable of performing three main functions: resource provisioning, virtual and physical resource monitoring and runtime control. Planning is performed considering the energy consumption of resources and the possibility of live-migration of VMs, using a bin packing genetic algorithm with multi-objective evaluation, while estimating server/virtual machine resource utilization and energy consumption using historical monitoring data. Monitoring refers to complex event stream processing, condition detection and action response. Runtime control refers to performing of control actions on both physical servers (start / stop) and on virtual machines (migrate), as well as to managing the virtual network resources (e.g. using OpenFlow - by configuring the network connectivity of VMs).

The planning system works in a control loop. The requests presented at the system's input are first passed to the allocator module for determining how the resources should be allocated on the physical infrastructure. The allocator uses the historical data from the monitoring module and the forecast of the system load for producing an allocation of VMs to hosts. The allocator produces several possible allocations, evaluating them using the multi-objective evaluator. The allocator might relocate the already allocated VMs if better system utilization can be achieved. The actual VM relocation is supposed to be performed by live migration. The control loop is closed by applying the allocation changes back to the existing infrastructure, followed by collecting of monitoring data and making it available to the load forecaster module.

Our current results include implementation of a control architecture for server and VM resources in the EU research project GEYSERS by using both the OpenNebula infrastructure manager and direct hypervisor access. We also integrated an energy estimation model for hosts and VMs based on a linear power model and created a framework for dynamic processing of monitoring events based on subscriptions for SLA monitoring and experimentation with VM planning algorithms (genetic algorithm with first fit heuristics and multi-objective evaluation).

Future work includes implementation of a SLA management architecture considering application control, infrastructure scaling and SLA monitoring, refining the cognitive algorithms for resource demand estimation by considering the repeating nature of cloud loads (VMs) and using it for better resource utilization prediction. Also, we plan to perform simulations of single and multi-domain resource allocation, considering the possible control operations and their impact on the application performance and overall energy efficiency of the system. Additionally, we will investigate other infrastructure management solutions (OpenStack, OpenFlow) in the context of Wireless Sensor Networks and Software Defined Networks.

Reliable Communication in Energy Efficient WSNs

Markus Anwander, Universität Bern
anwander@iam.unibe.ch

Wireless Sensor Networks (WSNs) are characterized by easy installation and adaptive self-organizing with no need for maintenance. The individual sensor nodes of WSNs are usually battery powered, equipped with sensing devices and use a low-power radio module for communication. The limited lifetime of batteries makes energy preserving mechanisms an important topic of current WSN research. Moreover, a WSN network stack has to offer reliable communication services to guarantee the functionality of the WSN applications. Unfortunately, reliability techniques require additional energy to recover packet loss. High traffic load involving intra/inter-flow interferences and congestion are especially challenging for energy aware reliability mechanisms. Additionally, WSN network stacks have to support standardized protocol headers to enable communication in heterogeneous WSNs. To solve these problems, we contribute a link layer with two novel mechanisms to support energy efficiency and reliability. Both mechanisms are able to work together with energy efficient and robust packet-oriented radio modules. The first mechanism is a novel traffic load measuring mechanism using traffic prediction. It predicts the expected incoming traffic load during periods of congestion. This enables selecting long sleeping periods to preserve energy while it simultaneously provides sufficient bandwidth during high traffic load or congestion. The second mechanism is a congestion detection mechanism that is able to identify and correctly handle intra-flow and inter-flow interferences as well as congestion. Our contributed WSN stack supports IEEE 802.15.4, IP, UDP and TCP to enhance the network connectivity and offers an energy efficient UDP based end-to-end reliability protocol. Moreover, we analyzed the impact of Forward Error Correction (FEC) codes on energy efficiency and reliability performance. The analyzed FEC codes are able to reduce the required local retransmission attempts but they are neither able to reduce the energy usage nor to enhance the reliability of our real world WSN network stack using a packet-oriented radio module. To evaluate our contributed WSN network stack we compared it to existing real world WSN network stacks. It uses less energy and shows a lower packet loss than all other stacks. Conducted experiments show that our WSN network stack is the most energy efficient reliable among all evaluated stacks. At the same time, our contributed link layer protocols show significantly higher throughput than the other evaluated protocols, i.e. up to 4 times higher throughput. This is due to the fact that our contributed link layer protocols are able to handle intra-flow and inter-flow interferences as well as congestion.

StreamHub: Scalable Content-Based Routing for Event Streams

Raphaël Barazzutti, Université de Neuchâtel
raphael.barazzutti@unine.ch

Publish/subscribe communication paradigm is becoming one of the most interesting choices for large information infrastructures that span over multiple physically distributed data centers (hybrid-clouds). Its success lies in its scaling properties as well as in the decoupled communications it relies on. By routing messages based on their content, publish/subscribe systems remove the need for applications to establish and maintain fixed communication channels between their components, thus yielding loosely coupled and highly flexible architectures. The problem of implementing efficient and scalable publish/subscribe mechanisms has been widely studied. Most of the proposed systems rely on networks of brokers organized according to the nature of the workload. In this presentation, we take a different approach and revisit the design of a high-throughput and low latency publish/subscribe system, emphasizing scalability and modularity. We present StreamHub, a publish/subscribe infrastructure that builds on top of an efficient event stream processing system, StreamMine. This engine provides support for fault-tolerance and replication.

StreamHub decouples the functionalities associated with scalability from those that deal with filtering. This allows the system to aim for the following goals : high scalability by supporting thousands of publishers and subscribers, high throughput by providing real-time matching of subscriptions against content of publications and low latency by timely delivering the publications to the subscribers.

StreamHub is built on top of StreamMine as a set of three stages (of operators) : access-point (AP), matchers (M) and exit-point (EP). Only matchers are stateful since they hold the subscriptions.

Since StreamHub aims to be used by applications written in other languages than C/C++, it provides a tool called DCCP which acts like an interface to the StreamHub engine. Subscriptions and publications can be sent to it through various serialization engines such as Google Protobuf and Apache Thrift, as well as several transport layers such as ZeroMQ and TCP sockets.

To achieve good performance StreamHub exploits the natural scalability of the content filtering operation by carefully partitioning publications or subscriptions and processing event streams in parallel, so that larger user populations or stricter throughput requirements can be supported by simply adding more machines to the system. StreamHub offers a feature called "clustering" which can be enabled at the access point level, it groups subscriptions with "similar properties" together in order to improve performance. The clustering algorithm has to be deterministic, two equal subscriptions have to be sent to the same matcher. Currently three algorithms are implemented : hashbased, K-means and event space partitioning.

The scalability and performance of StreamHub are evaluated by using a prototype implementation on a 48-node cluster.

Fault Tolerance Based on POP-C++

Jianping Chen, École d'ingénieurs et d'architectes de Fribourg
Jianping.Chen@edu.hefr.ch

With the development of high performance computing, the scale of distributed system becomes larger and larger. But meanwhile, with the nodes or processes increasing, the probability of failure in distributed system also increases. To tolerance failures in distributed system becomes more and more important.

In the field of fault tolerance, there exists several kinds of techniques, such as replication(active replication or passive replication), checksum, checkpoint, N-version programming, SWIFT and message logging. All of these techniques are based on the conception of redundancy. Replication as literature shows, it makes several replicas for sever or data, and often be used in type of failure, server crash down. Checksum is a technique which is originated from fault tolerance for hardware, especially for the memory. According to add some redundant data make verify the correctness of the computing. SWIFT is short for software implemented fault tolerance, and it is a software-only, transient-fault-detection technique. SWIFT efficiently manages redundancy by reclaiming unused instruction-level resources present during the execution of most programs. But among these techniques, checkpointing and message logging are most frequently used in distributed system thanks to their lower cost of hardware and relatively simple protocol.

Checkpointing is a very common technique which is used to make fault tolerance in distributed system. By using checkpointing, process will be recorded including data, stack, register and environment variations. Through designing appropriate protocol for checkpointing, checkpointing is divided into two categories, coordinated checkpointing and uncoordinated checkpointing. Compare to coordinated checkpointing, uncoordinated checkpointing can make checkpoint independently for different processes and will not block the execution of processes. Although uncoordinated checkpointing has a lot of advantages, but uncoordinated checkpointing may cause orphan process and domino effect.So usually, uncoordinated checkpointing is used to cooperate with message logging.

Message logging is a common method of building a system that can tolerate processes crash failures.For determinant event in the distributed system, checkpointing can do well during process recovery. But for non-determinant event, the results can not be redelivered by only using checkpointing technique. Message logging provides the ability to solve this problem of non-determinant event by recording the results of events which the process made before it crash.

But traditional message logging will log messages of each process. This will incur recording a lot of data, make a lot of memory or storage overhead. To alleviate this overhead is a research field of message logging.We thought it is not necessary to record message for each process. We can enlarge the granularity of message logging. Here we propose a new logging strategy, logging messages by grouping the processes. In POP-C++, object can be created remotely and works as a process. All the objects register the information on the main program. Due to this feature of POP-C++, we can group the parallel objects by functions, making partitions for the objects.We can analyze the relationship between the objects during the compiling time of POP-C++. Then group objects which will finish the same function to a partition. This partition will work as a unit, message logging strategy will only log the messages which are delivered as input or output of the whole partition, but not log the messages transferring inner partition. We call this method as partition fault tolerance.

A Unified Technique to QoS Guaranteed Scheduling in LTE-Advanced Networks – Emphasis on System Throughput and User Fairness

Ioan S. Comsa, École d'ingénieurs et d'architectes de Fribourg
Ioan.Comsa@beds.ac.uk

The growing demands for multimedia services such (e.g. time gaming, conversational video, file transfers) have enforced 3rd Generation Partnership Project (3GPP) to work with Long Term Evolution – Advanced (LTE-A), a novel radio access technique based on Orthogonal Frequency Division Multiple Access (OFDMA) multi-user modulation scheme. However, the new technology is not enough to face the problem of the increase of data traffic due to the limited nature of radio resources. Therefore, an advanced management of radio resources is strongly recommended.

The packet scheduling is a particular sub-module which is used to assign resources to each user in order to deliver services in an efficient way. Packets are scheduled in every Transmission Time Interval (TTI) which is a time window used for the user requests delivery on both uplink and downlink sides. In LTE-Advanced, the TTI duration is 1 ms revealing the sensitivity of the scheduling process and its impact on the resource management. Moreover, the scheduling technique is based on the scheduling discipline which is used to give different priorities of serving the user requests. The scheduling rules are various because of the different conditions which are taken into account. Therefore, each scheduling rule has its own property and behaves in different way depending on the scheduling target which we want to achieve.

In general, the packet scheduling performance is analyzed in terms of different metrics such as: system throughput, user fairness, Quality of Service (QoS) guaranties, system stability, power efficiency and system complexity. By applying one scheduling rule for the entire scheduling process one of the mentioned target will be satisfied with the price of harming the other ones. For instance, the rule called maximum throughput aims to maximize the system throughput but is unfair for users with worst channel quality. On the other side, round robin is the best rule from the user fairness point of view, indicating a poor performances on system throughput. Largest-weighted-average-delay-first (LWADF) assumes to guarantee the delay constraints for real time services, harming the system throughput and the user fairness for the other types of services.

An efficient scheduler should be able to allocate resources in order to respect the satisfaction level in which as many performance metrics as possible to remain high enough. The idea is to use different scheduling rules for each TTI instead of the single one for the whole process. Each rule has to be applied for each time instant in order to maximize the objective function.

In this paper, a novel technique is used for the particular case where different levels of trade-off between user fairness and system throughput can be achieved based on different classes of users. The trade-off concept between throughput and fairness plays an important role in special for the best effort traffic where the operator is responsible to increase the system throughput as much as possible maintaining different level of fairness.

By imposing a fixed threshold for throughput or fairness performance makes the scheduler inflexible when different trade-off levels are required by the environment. The proposed objective function is composed by the percent of users which are located in three different classes depending on the Channel Quality Indicator (CQI).

A number of 40 scheduling rules are used respecting the Pareto optimality from the fairness and throughput point of view. The policies adoption and refinement is done by using neural

Q learning algorithm which is based on the dynamic programming and Markovian processes in order to develop different policies in the long-term running purpose. The state space is defined by percentage of users located in different classes, the Average Jain Fairness Index (AJFI) and the Average Normalized System Throughput (ANST). By applying one scheduling rule the system will move from one state to other one. In the exploration stage, the rule selection is based on the Boltzmann probability distribution.

In order to handle a large state-action space in our scheduler, we use the neural network which permits to interpolate Q values for the unvisited state-action pairs.

Six policies are analyzed based on different fairness-throughput trade-off levels depending on the percentage of users for each class. Simulation results indicate that the novel proposed method outperforms the existing scheduling techniques by maximizing the system throughput, user fairness, cell spectral efficiency and cell-edge spectral efficiency. Moreover, the system achieves a desired throughput-fairness trade-off and a global satisfaction for different classes of users.

Multimedia Intrusion Detection in Multi-tier Wireless Multimedia Sensor Networks

Denis do Rosàrio, Universität Bern
rosario@iam.unibe.ch

The proliferation of multimedia applications and the demand for multimedia services in Wireless Sensor Network (WSN) [1] are creating a new multimedia sensor-based era, and have fostered the development of Wireless Multimedia Sensor Networks (WMSNs) [2].

WMSNs promise a wide scope of potential applications in both civilian and military areas, which require visual and audio information. In this work, we focus on the application of multi-tier WMSNs with video support for an intrusion detection scenario. In such application, real-time video content has the potential to enhance the level of collected information as well as, to detect and to monitor intruders. However, the video sequence should be delivered to the end user with an acceptable video quality level from the user perspective, enabling intrusion detection. It has been shown by many works that a multi-tier architecture offers considerable advantages with respect to single-tier architectures in terms of lower energy consumption, better scalability, lower loss, higher functionality, and better reliability. In this work, we propose a multi-tier WMSN architecture composed of only three tiers.

The scalar sensor node (tier-1) are performing intrusion detection, e.g. using vibration sensors. The camera nodes (tier-2) can be woken up on-demand to retrieve real time video of the intruder that has been detected previously by the lower tier (scalar sensor), and send the video stream to the base station. The mobile camera node (tier-3) should be woken up in a case of intruder detection. Then, it will move to the area of the event, retrieve multimedia content from a defined region and send it to the BS.

For low-power communication, the constraints of sensor nodes will increase the effects of wireless channel errors. Thus, some error control schemes for multimedia communication over multi-hop WMSNs are needed. Application-level Forward Error Correction (FEC) can be applied as error control scheme for handling loss in real-time communication. FEC adds h redundant packets to n original source packets on application layer to enable recovery of lost packets. The recovered data can be used to reconstruct a lost frame, and thus improve video quality.

A multimedia intruder detection application requires high video quality from the user perspective, scalability, energy-efficiency and low network overhead. However, existing works do not take into account FEC mechanisms that consider the frame importance to create redundant packets. This can improve the video quality without increasing the overhead and energy consumption. Additionally, these proposals do not use multi-tier architectures to provide energy-efficiency and scalability.

In this context, this work proposes: (i) a Quality of Experience (QoE)-aware FEC mechanism in WMSNs for an intrusion detection application using a multi-tier architecture; and (ii) a reliable handoff mechanism for mobile camera nodes sending multimedia packets through static nodes.

The QoE-aware FEC mechanism considers the frame importance and its impact from the user point-of-view to create redundancy. The loss of I-frames causes more distortion than the loss of P- and B-frames from the user perspective. Additionally, the loss of P-frames at the beginning of a GOP causes more video distortion than at the end of a GOP.

In this context, the QoE-aware FEC mechanism applies the following rule to create packet redundancy. I-frames urgently need redundancy, since their loss will cause more video distortion. On the other hand, B-frames do not need redundant packets, since they are not used as a reference for other frames. The redundancy for P-frames can be applied based on their position within the GOP, i.e. the first 50% of P-frames of each GOP will have their packets encoded using redundancy and the last 50% of P-frames of each GOP are sent without redundancy.

The mobile camera nodes should send the multimedia packets to the base station through the static nodes. This is a typical scenario for a handoff process, where the mobile nodes should select the best available node (access point) to send the multimedia data packets. However, there are several challenges to perform a reliable handoff, due to the fact that WMSNs have severely constrained resources. Furthermore, low-power links have short coverage and higher variability.

Our proposals aim to reduce the overhead, delay and number of handoffs, while keeping the video with high quality from the user perspective. Expected results will bring many benefits in a resource-constrained system.

References

[1] Yick J., Mukherjee B., Ghosal D., Wireless sensor network survey, Computer networks, vol. 52, no. 12, pp. 2292-2330, 2008.

[2] Almalkawi I., Guerrero Zapata M., Al-Karaki J., Morillo-Pozo J., Wireless Multimedia Sensor Networks: Current Trends and Future Directions, Sensors, vol. 10, no. 7, pp. 6662 6717, 2010.

Analysis of Anomalies in the Internet Traffic

Veronica del Carmen Estrada Galinanes, Université de Neuchâtel
estrada.veronica@gmail.com

1 Background

Intrusion Detection Systems (IDSs) are used to detect exploits or other computer attacks that raise alarms. Some IDSs are able to detect anomalous events that indicate abnormal activity in the network. On one hand, abnormal behaviors may be legitimate, e.g., misinterpreted protocols or malfunctioning network equipment, but on the other hand they may introduce vulnerabilities and facilitate future intrusions or they could be even a maliciously crafted packet with the purpose of evading the security monitors.

Security administrators are burdened with the analysis of security logs and frequently overlook anomalous events. Recently, researchers have been enthusiastic about applying machine learning (ML) algorithms to automatic or semi-automatic learning in security logs. However, there is low acceptance of these techniques in the operational real-world network environment [1]. A possible explanation is that ML methods are not employed effectively; a considerable portion of the research on ML applied to intrusion detection employs the KDD benchmark, an outdated data set based on a simulated network with a limited environment. It contains 5 millions of connection described using a 41-dimension feature vector with a label indicating normal or attack type. KDD attracts significant attention from researchers due to its well-defined and readily accessible nature. In fact, the use of a pre-processed dataset such KDD may reduce up to 80% of the project time (rule of thumb). However, the careless overuse of KDD is a disgrace to this field since the benchmark is full of deficiencies and limitations exposed by several experts immediately after its release.

Alternative datasets and public available raw data were captured mainly with the purpose of traffic engineering. Thus, they are not oriented to our security needs, for instance, attacks are not classified and full payload is not always available.

2 Aim

Considering that we want to achieve more efficient use of ML in the future and the quality of the sample to train algorithms is critical, at this stage, we give a priority to gaining knowledge from real traffic behavior. In order to build a framework of reference for understanding anomalous behavior, our proposal is auditing real, up-to-date and large amount of data as part of the cycle of pre-processing captured network packets to select reliable features for future ML benchmarks.

3 Method

The characterization of anomalies in a large network environment is a topic rarely investigated. Questions and problems arise in data pre-processing and are not anywhere documented. Our observations are based on the analysis of anomalies in the Internet traffic observed at the University of Tokyo campus network gateway. We analyze 6.5 TB of compressed binary tcpdump data representing 12 hours of network traffic. The analysis covers a wide aspect of anomalous behaviors. To achieve this goal, we examined patterns in more than 60 millions of connections, which are segregated in incoming and outgoing traffic. We aggregated data to analyze frequent episodes, exposing some DoS and probing attacks by observing SYN and fragments among other indicators. Then, we include observations on activity not consider in the KDD benchmark.

The main tool for detecting attacks and anomalies is BRO IDS, a powerful network analysis framework. The reason why we used BRO as the principal analyzer is because an earlier version of the system was used to generate the KDD benchmark. More deep analysis was complemented with outputs from other popular network tools such Argus and Wireshark. A portion of the observations was verified with Snort IDS.

4 Results

Our major contributions can be summarized in: 1) reporting the anomalies observed in real, up-to-date traffic from a large academic network environment, and documenting problems in research that may lead to wrong results due to misinterpretations of data or misconfigurations in software; 2) assessing the quality of data by analyzing the potential and the real problems in the capture process. In particular, we found a substantial amount of anomalous behavior not considered in KDD related with bugs in applications, abnormal traffic generated in overlay networks, hidden services running on non-standard ports, protocol violations and techniques for evading the security barriers.

5 Conclusions

When using a pre-processed dataset, researchers do not have the opportunity to face the limitations that appear when real data is analyzed. Besides, in the cases that employ KDD data, studies are limited to 41 features that describe the traffic generated by user connections with the erroneous assumption that they can describe real, actual Internet traffic.

We present diverse pathologies found at different levels of the network stack covering a wide variety of cases. Using the KDD dataset as a reference, we identified 5 behaviors that are missing from KDD dataset and could be defined as new features or can be used to rule out data that may distort statistics and consequently reduce the detection power of machine learning algorithms.

6 Limitations

To achieve our goal, there was a huge amount of manual processes. Each environment has unique characteristics and researchers need to learn them before using data for metrics.

The difficulties in capture process should not be underestimated. Flaws usually appear in datasets and the way we handle them may impact on measurements. Artifacts introduced by the capture machine should be detected in the early stages of the research. In this work, we assessed the quality of the captured data and noticed two deviations: malformed packets caused by the capture process and non-compliance 3-way handshake.

Finally, the detection capacity of intrusion detection is highly influenced by the system configuration. We documented problems that may lead to wrong results due to misinterpretations of data or misconfigurations in software (IDS configuration “by default”). Participants with diverse backgrounds do research on network anomalies. Thus, our approach is trying to understand better the nature of common anomalies found in current operational networks to benefit researchers coming from network, information security and machine learning areas.

Software Defined Radio and USRP Synchronization with GPS

Zan Li, Universität Bern
li@iam.unibe.ch

1 In3DGuide Project Introduction

Localization has facilitated people's daily life through the use of Global Positioning System (GPS) devices. However, GPS signals cannot be used in indoor environments where localization is necessary for some applications, such as pedestrian navigation, locating firefighters etc.. Therefore, some research has already been done to realize localization by some other signals, such as WiFi and Bluetooth signals. In our project (In3DGuide), the GSM signal, which is more common used than the others, is aimed to be applied to localization.

2 Software Defined Radio

A Software Defined Radio (SDR) system is a radio communication system where components that have been typically implemented in hardware are instead implemented by means of software. A SDR system is divided into three blocks. First, the Radio Frequency (RF) frontend of the hardware serves as an interface to the analog RF domain. Then, the intelligence of the hardware part is implemented, forming the interface between the digital and the analog signal. Finally, nearly all of signal processing parts including modulation, demodulation etc. are implemented by software which can be installed on personal computer or embedded devices.

In our project, the hardware, which includes the first two blocks, is the USRP E110 device. The USRP Hardware Driver (UHD) provides a host driver and Application Programming Interface (API) for the USRP. A Gumstix Overo Single-Board Computer (SBC), where the GNU Radio software is installed and run, is integrated in the USRP E110 device.

3 USRP Synchronization with GPS

3.1 PLL Synthesizers

A Phase Locked Loop (PLL) frequency synthesizer is a feedback control system. The phases of two input signals are compared in the phase comparator and produce an error signal which is proportional to the difference between their phases. The error signal then passes the low pass filter, and then drives a Voltage-Controlled Oscillator (VCO) to create an output frequency. The output frequency is fed through a frequency divider back to the input of the system, producing a negative feedback loop. If the output frequency drifts, the phase error signal will increase, driving the frequency in the opposite direction to reduce the error. Thus the output frequency is locked to the frequency at the other input. This other input is called the reference and is usually derived from a crystal oscillator, which is very stable in frequency. [2]

There is a 2.5 ppm Temperature Compensated Crystal Oscillators (TCXO) frequency reference in the USRP E110 device. To increase the stability of the sampling frequency, an external reference clock can be selected. In our project, the GPS Disciplined Oscillator (GPSDO) reference clock with 0.01 ppm and 10 MHz is applied. [1,5]

3.2 Time Synchronization

GPS has provided a way to achieve precise time synchronization. It has, in effect, offered the services of precise atomic clocks for free to anyone around the globe with a suitable GPS receiver and a clear view of the sky. Many GPS receivers provide a timing pulse, the so-called “one Pulse Per Second (1PPS)” signal. This pulse normally has a rising edge aligned with the GPS second, and can be used to discipline local clocks to maintain synchronization with Coordinated Universal Time (UTC). The characterization of the 1PPS timing pulse from GPS receivers can provide useful information to anyone contemplating using GPS for timing applications [6].

As introduced before, the Ettus Research GPSDO Kit for USRP N200 and E100 series was applied in our work. It provides a PPS signal with an accuracy of ±50ns to UTC time. The PPS signal is applied to trigger the time registers in the USRP devices to start from the GPS time plus one. Therefore, the times in all USRP devices can be synchronized to the GPS time.

4 Two Experiments for USRP Synchronization

4.1 Setup and Preliminary Results of the First Experiment

According to the method which is mentioned in section 3, we used the UHD to setup two USRP devices to check if they can work in the same GPS time. The key steps are as follows. First set the clock and time source to the external source (GPS). Then time registers in the USRP devices are triggered by the PPS to start from GPS time plus one. Finally, after the modification in the example file (rx_timed_samples) in the UHD, rebuild it and run to test.

According to preliminary results, if the times in two USRP devices are synchronized to the PPS and start receiving at the same time, the timestamps of the samples from both USRP devices are exactly the same. That means that the two USRP devices are synchronized well. However, there are still two problems in the synchronization. The first one is that there may be different delays in the RF frontends and Analog to Digital Converters (ADC) in two USRP receivers. The second one is that the resolution of the clock (52MHz) is around 19ns which may be not good enough for accurate localization.

4.2 Second Experiment

In order to check the two problems mentioned before, we plan to perform another experiment. For the second experiment, we aim to use one USRP device to transmit some packets to two USRP receivers. The distances between the transmitter and the two receivers should be set to the same distance value. Then, we can read the timestamps from two receivers to check if the same packet can get the same timestamp in both receivers. This experiment is still work in progress.

References

[1] E110 Product Details. UE110-KIT. https://www.ettus.com/product/details/.

[2] Frequency Synthesizer. http://en.wikipedia.org/wiki/Frequency_synthesizer. Accessed: May 2012.

[3] GPSDO Data Sheet. Technical report, Ettus Research.

[4] GPSDO-KIT Product details. https://www.ettus.com/product/details/GPSDO-KIT.

[5] USRP E100/E110 Embedded Series Data Sheet. Technical report, Ettus Research.

[6] P.J. Mumford. Relative Timing Characteristics of the One Pulse Per Second (1PPS) Output Pulse of Three GPS Receivers. The 6th International Symposium on Satellite Navigation Technology Including Mobile Positioning & Location Services, July 2003.

Supporting Time-Based QoS Requirements in Software Transactional Memory

Walther Maldonado, Université de Neuchâtel
walther.maldonado@unine.ch

Multi-core processors are becoming the norm in today's computing environment, from desktops to laptops and cellphones to tablets. In order to successfully exploit the performance of these processors, programs must make effective use of multi-threading. However,the traditional approach of locks has many setbacks which prevent them from being widely used in parallel applications.

Software Transactional Memory (STM) is an optimistic concurrency control mechanism for simplifying parallel programming. By tracking the state of memory during reads and writes, it can treat code-blocks as atomic transactions, and abort and retry conflicting transactions as needed to preserve the consistency of the system.

Still, its applicability to reactive applications which have time constraints on specific tasks has not been well tested. These applications are referred to as reactive applications, as they have a required response time associated to certain operations.

We propose supporting these type of applications by allowing transactional blocks to be annotated by the programmer with required execution completion times in the form of *deadlines* and quality of service (*QoS*). The deadline specifies the amount of time under which the operation must be completed, while the QoS specifies how much we may degrade while still delivering acceptable performance (e.g.: a QoS of 99

Meeting the deadline is done through two mechanisms: time estimation and execution mode switching. Past execution times are taken as samples to estimate the duration for future transactions. Using Vitter's reservoir algorithm, a subset representative of past executions is kept, and the expected duration L is taken by using the nth percentile from the sorted reservoir.

The execution modes have decreasing levels of optimism while increasing commit likelihood, in this way the progress of other threads is not slowed unless it is required to meet a deadline on time.

The execution modes are as follow:

- Optimistic: Offers the highest degree of parallelism when contention is low. However, reads are not tracked, and thus *Read, then Write* conflicts are not detected until commit time.
- Visible Reads: Tracks memory reads as well as writes. Detects *Read, then Write* conflicts, though it has higher overhead.
- Irrevocable: Ensures commit by acquiring a global lock which prevents other Read/Write transactions from committing. However, it can serialize execution if used frequently.

To deal with concurrent deadlines, we propose different conflict resolution policies to handle QoS degradation fairly between the threads and avoid starvation. Fair degradation consists in having a similar performance (ratio of deadlines met, *hit-rate*) between the different threads, as well as keeping the time by which a deadline is missed consistent (*overflow time*).

The concurrent-aware contention managers are as follow:

- Basic: Time to deadline is the criteria used to resolve conflicts (i.e.: there's no fairness considered, just throughput)

- Fair: The current hit-rate of the thread is used.
- Cycles: The amount of overflow-time is used.
- Compound: If both deadlines have been missed already, priority goes to the thread with the most overflow-time. Otherwise, time to deadline is used.

Additionally, to handle concurrent deadlines that compete to use the irrevocable mode at a time, a queue was implemented to ensure ordered execution. Two types of queue were devised: ordered by the time remaining to deadline of each thread, or ordered by the overflow time of each thread.

In order to limit the excessive use of irrevocable executions (which serializes the system), different *thresholds* were experimented with. The threshold specifies the number of concurrent deadlines under which irrevocable mode may still be considered (e.g.: if the threshold is four, when there are five or more concurrent deadlines, irrevocable will not be used).

Our contribution consists of an extension of tinySTM to gather duration statistics and switch execution modes, as well the various deadline-oriented contention managers. The experimental results show that our methods considerably increase the likelihood of meeting the deadlines on time, with fair QoS degradation for cases with very high contention. For concurrent deadlines, the overall best results where achieved by combining the used criteria (e.g.: hit-rate based contention manager with an overflow time based irrevocable queue). The thresholds had very limited impact: they only improved performance for combinations that had much worse performance than the others.

Accounting in Wireless Mesh Networks

Alexey Monakhov, Université de Neuchâtel
alexey.monakhov@unine.ch

1 Introduction

A Wireless Mesh Network (WMN) may extend a network converge in many situations including network access on a campus, at a conference, during environmental research, and other kinds of temporary activities wherein a wired infrastructure is not available. It is a quite common for different organizations to cooperate in building and maintaining a network infrastructure, because it may be beneficial in terms of cost saving and independence from third party organizations. In the ongoing research, we propose traffic based accounting in shared WMNs, wherein different parts of the network belong to different organizations. We provide organizing schemes and design requirements of our accounting system as a part of the whole network architecture. This approach was implemented in the A4-Mesh project by creating accounting infrastructure, which allows us to have the information on a contribution of every node in traffic forwarding and traffic generating. This information is supposed to be used later on for billing, network management, and planning. Additionally, we propose reliable accounting, which allows us to verify accounting information gathered from all nodes. It should work as an extension of the basic accounting architecture.

2 Goals of the Research

In the ongoing research, we have the following goals: design and implementation an accounting system suitable for WMNs with existed authorization and authentication (AA) infrastructure operating in the federated scenario introduce and implement our reliable accounting

3 Existing Solutions and Challenges

service oriented scenario. This scenario implies that user is connected to the provider in a one hop and traffic is not counted for intermediate nodes, only generated traffic by user is subject to count without distinguishing who transferred it. time based accounting - accounting is considered only as time based tracking procedure, i.e. we track time when user started and finished using resources. In a case when we need to have users always connected to the network this model is not applicable. We can summarize requirements for targeted accounting architecture: it is necessary to have detailed information about contribution in traffic forwarding of every intermediate node modular architecture of the system which allows us to make modification in their modules (for example using any AA infrastructure which is not coupled with accounting system).

4 Proposed Accounting Architecture

We present notion of accounting as a process which consists of 3 stages:

1. Resource consumption monitoring stage where we use Netflow based solution to track traffic which was generated and forwarded by particular node.

2. Linking during this stage we associate IP address derived from the stage 1 and a UserID obtained from the authentication service.

3. Further data processing (aggregation of thousands records and storing data into the database). Difficulty of the aggregation task is that on the one hand we should keep size of the database in some borders on the other hand we need our data to have sufficient level of granularity.

5 Reliable Accounting

Based on analysis of existed problems and potential improvements of accounting system we concluded that the problem of reliability of accounting information is one of the most important. We assume that network participants can claim incorrect information about traffic which they forwarded or generated during network operation. For example we can provide the following scenario of cheating: node claims more traffic forwarded than it transferred indeed node overhears the traffic and claims that it participated in the traffic transmission node changes the routing, i.e., it involves more nodes to the transmission than it is required node to which a client is connected generates additional packets on behalf of the client

6 Existing Approaches and Drawbacks

We consider two existed approaches – micro-payments which implements the idea that every node has virtual money and for any data transferring activity node should get some amount of this virtual money first. This scheme requires additional infrastructure to establish this micro-payments infrastructure (brokers, authorization) which leads to complexity of the system. Also if node does not have credits then operation of the network will degradate if other node will not transfer packets of such a node (which is supposed to be according to this approach). confirmation of every packet node transferred. The idea is straight forward – every node confirms every packet it transferred using some kind of encryption to be sure that node can confirm packets which it really observed and nobody else can overhear it. Such a system have quite complex protocol for confirmation of every accounting message which leads to computational and traffic overhead about 10-20 per cent.

The plan is to use the second approach but propagate it for flow based approach (which will reduce traffic overhead) and to introduce scheme where neighbor will confirm contribution of each other – it will improve reliability of the accounting system.

7 Suggested Architecture

To implement this idea we suggest including additional information into the NetFlow records about every traffic flow observed by the node - “from whom” - node writes down node from which the packet was received “to whom” - node writes down node to which the packet was transferred hashFunction(flow) - node calculates a fingerprint for the flow of packets it observed On the gateway (GW) we perform verification procedure to check consistency of information claimed by every node with assumption that GW itself is trusted. This procedure looks like constructing a chain which is composed out of trusted (i.e., checked) nodes starting from the

GW. If we find an inconsistency on some stage it means that the chain is broken and associated nodes cheat with accounting information. To identify the cheater we assume that it can be only one cheater in the chain and if we start to build similar chain starting form the source node then this chain also will be broken and we will find a cheater. Potential problems needs to be solved - correctness of this procedure and verifying that we have not too strong assumption and that cheating case which we prevent by this procedure is not trivial – it requires to develop adequate description of all possible cheating cases.

8 Ongoing Problems, Future Work and Possible Extensions

In this scheme we depend on encryption scheme which is used and existence of secured channel between every node and the GW. It can lead to difficulties in deployment of such an accounting system and problems with configuration. So the next stage would be to design reliable procedure which does not require such an encryption. Another challenge is computation overhead which we have on the GW, thinking about improvement we can try to introduce some ideas about distributed accounting.

Thrifty Privacy: Efficient Support for Privacy-Preserving Publish/Subscribe

Emanuel Onica, Université de Neuchâtel
emanuel.onica@unine.ch

Content-based publish/subscribe is an appealing paradigm for building large-scale distributed applications. The typical setting for such systems involves a broker overlay meant to store subscriptions registered by subscribers. These are matched with publications submitted by publishers, which are routed afterwards towards the interested subscribers. Various use cases have been described in the literature fit in such scenarios. The most often referenced is the stock exchange scenario in which the subscribers are represented by investors registering stock queries and the publishers are financial agencies publishing stock information.

Privacy becomes a major concern in such systems when these are deployed over untrusted administrative domains. The basic action of routing data based on the matching between the publications sent in the system and the stored subscriptions can lead to sensitive information leakage. In the mentioned use case somebody can imagine an attacker exploiting an infrastructure deployed over a public cloud through virtual machine collocation attacks that can lead to leak the information stored at the brokers level. This could be used afterwards to influence the financial market. Such attacks can be avoided by the means of privacy-preserving filtering. These are more exactly represented by schemes that allow encrypting the messages (publications and subscriptions) sent into the system in such a manner that the untrusted brokers can perform the matching and subsequent routing, but can not access the actual information in the messages. Therefore, an attacker does not get any access to sensitive information.

Unfortunately such approaches tend to cause a high performance overhead and restrict also the possibilities of classical optimization for content-based filtering such as per-attribute containment. In particular in our study we make use of a solution inspired from the database security area [1] that was further adapted for the publish/subscribe setting: Asymmetric Scalar-product Preserving Encryption (ASPE) [2]. Even if the communication overhead is acceptable and the scheme design is simple to be applied in a practical setting, like also in other cases the matching complexity is costly. This goes up to quadratic in the number of evaluated constraints in a subscriptions which is a significant drawback compared to simple plaintext matching.

Our work [3] consists in a novel mechanism that reduces the costs inferred by the usage of encrypted matching operators. This mechanism is implemented as a prefiltering stage that takes place at the broker level before the encrypted matching. The main idea is to avoid the costly encrypted matching operation by determining in advance of it if the subscriptions do not match the publications.

The prefiltering stage is based on evaluating equality constraints. These are common in a variety of scenarios. For instance, in the stock quotes mentioned case the subscriptions will most likely include a equality constraint on the stock quote symbol. The proposed mechanism is based on encoding these constraints in a Bloom filter representation that is added to the subscription. Also, the attribute values in the publications are encoded in similar Bloom filter representations. The encrypted messages are sent into the broker infrastructure augmented in this manner. The basic principle of prefiltering applied at the broker level is to determine if a publication does not match a subscription by evaluating if the bits in the subscription Bloom filter are not found among the bits in the publication Bloom filter. If this is the case, the certain result is a *no-match* and the encrypted matching operation can be avoided. If the opposite happens, all the bits in the subscription Bloom filter are among the bits in the publication Bloom filter, then the encrypted matching must be executed in order to avoid a false positive *match* result. This basic mechanism of prefiltering is further enhanced by exploiting containment relations between the subscriptions. This reduces the number of possible matching checks. Details about the prefiltering stage are presented in [3].

One important part of our work consists in analyzing if and how much sensitive information is leaked through the above mentioned Bloom filter augmentation brought to the encrypted messages. The problem comes from the fact that normally the encryption scheme used offers containment determination in an optional manner. This could be for instance disabled if it is considered that an attacker can make use of such information. However, the added Bloom representation offers partial containment information. Therefore we strengthen our scheme by randomly modifying the Bloom filter representation in a manner that still preserves our basic principle of *no-match* determination but does not offer containment information. In brief this consists in removing bits from the subscription representation (or/and adding bits to the publication representation). Our privacy analysis shows that deriving accurate containment information from this representation is highly unlikely. An extended description of our results is available in [3].

Besides the highly encouraging performance results obtained through our mechanism, the prefiltering stage comes also as a viable solution from an engineering point of view. The described enhancements were tested in relation with ASPE as encrypted matching, but can be potentially applied to any other encryption matching scheme due to their decoupled nature.

References

[1] W.K. Wong et al., *Secure kNN Computation on Encrypted Databases.* SIGMOD 2009

[2] S. Choi et al., *A Privacy Enhancing Content Based Publish Subscribe System Using Scalar Product Preserving Transformations*, DEXA 2010

[3] R. Barazzutti et. al., *Thrifty Privacy: Efficient Support for Privacy-Preserving Publish/Subscribe*, DEBS 2012

Integrating IoT and Enterprise IT: On the Need of Semantic Descriptions for Interoperability

Matthias Thoma, SAP Research, Universität Bern
matthias.thoma@sap.com

The enterprise IT world nowadays focuses mainly on service oriented architecture (SOA) and service orchestration and starts to adapt service choreography. Business processes are described in modeling languages like BPMN or BPEL. Business process decomposition then decomposes the process into distributed process steps, where some of them can be executed even on physical devices like sensors or a sensor network gateway.

The integration of sensor networks as first class SOA citizens, and thus allowing easy integration into ERP systems is a very recent research topic. Software development for sensor networks necessitates a good knowledge of the underlying technical details, which makes business integration a cumbersome activity. There is an obvious gap in enterprise integration, which needs to be addressed, between the enterprise SOA world with its general purpose execution engines and process modeling on the one hand, and the sensor network world with its custom (often manually tuned) embedded software on the other hand. Furthermore, additional aspects, like evolution and timesharing, quality of service and service level agreements need to be addressed as well.

The following five drivers in the integration of sensor nets into enterprise IT system need to be addressed:

1. Service Characteristics
   For the integration of a service into an enterprise system it is necessary to have a service endpoint and at least a technical description on how to invoke it. Additionally, it is desired to have also non-functional aspects described, which the execution engine can take into consideration in the binding and execution phase of a business process.

2. Service Discovery
   In a traditional SOA environment a service registry or repository is sufficient for discovering services within an enterprise. For those sensor networks, which are more or less static in nature, where most of the business logic is performed in the enterprise backend system, or if only limited business logic is executed on the nodes, a repository approach is sufficient as well. Nonetheless, in ad-hoc or self-organizing scenarios, where a lot of business logic is executed on the nodes and a backend system might not even exist, self-description of services is a must. In an industrial setting there is often the problem that one or more sensor nodes join a different enterprise context, depending on their location. In transportation scenarios, for example, a sensor network would monitor the goods along the complete supply chain. In case of food it could for example monitor humidity and temperature.

3. Specialized Application Fields
   We think that there is no general solution for interacting with real-world entities. While a service oriented approach with web services and SOAP or REST interfaces has worked pretty well in large enterprise IT systems, this is not necessarily the case in other application fields. Sensor network research has so far discovered a lot of different solutions for the specific needs of different applications and deployments. There exist a lot of specialized protocols for WSNs with different advantages and disadvantages (regarding reliability or bandwidth constraints for example), which are tailored for special usages.

4. Constrained Resources
   It is essential for the wide adoption of sensing technology in the enterprise IT that the related technologies come at a low cost. Therefore, we are dealing with devices, which are constrained in terms of memory, computation and communication. In an industrial setting the usage of constrained devices is desired and often enforced to reduce the total cost.

5. Reconfigurability and Programmability
   As more sensors are being deployed by enterprises, the evolution, shared use and reuse of already deployed sensor networks will play a crucial role. In a typical enterprise IT system a sensor network will be used for one or multiple tasks for some time. Changing requirements and cost pressure will lead to the need of a constant reconfiguration and shared use of resources. Applications, which time share a sensor network, might need to reflash a node to perform its task, due to the fact that sensor nets are usually memory constrained. Therefore, it is essential that the possible services and their requirements are properly defined.

6. Business integration
   Most work in the area of sensor networks focus on more or less technical aspects, only. Nonetheless, it is necessary to have a link from the sensor network to business aspects, like pricing and service level agreements. In an intelligent container context, it is then possible to run small business processes on the sensor nodes, for example, calculating the final price the customer has to pay based on SLA and pricing models stored on the sensor nodes.

The goal of the research is to utilize semantic technologies, namely SAPs USDL (Unified Service Description Language) based on the Resource Description Format (RDF), and the Linked Data Approach to accomplish a domain-aware integration of sensor networks and sensor nodes into a SOA environment. Additionally, the six drivers will be tackled in the following ways:

Service Characteristics, Service Discovery and Specialized Application Fields: A new way to integrate sensor networks, and other building blocks of the Internet of Things, in a service oriented way into enterprise IT systems is to use semantic technologies. Currently, there is no high-level service description language tailored towards embedded devices. Linked USDL will be extended to allow domain-specific annotations.

Constrained Resources: An adoption of Linked USDL for the description of services provided via small embedded devices will be done. This includes the technical challenges of implementing and accessing RDF servers on sensor nodes, compressing RDF files, discovery mechanisms and handling of binary software code.

Reconfigurability, Time-Sharing and Programmability: We foresee the need for landscape management, which makes evolution and time-sharing of corporate sensor networks possible. At this point, one idea is to reuse what has been done in cloud computing. There are quite some similarities between cloud computing and sensor networks regarding reconfigurability, time-sharing and programmability.

Business Integration: SLA management could be done the same way it is currently done for in cloud computing contexts. USDL already contains a pricing module which could be adopted for the the use within a sensor service description.

Decentralized, Scalable and Trusted Storage & Indexing

José Valerio, Université de Neuchâtel
jose.valerio@unine.ch

Abstract

This doctoral thesis is focused on large-scale storage systems. For such systems, SQL-like solutions (RDBMS) are not suitable because of their lack of scalability. Alternative designs (i.e. NoSQL) are a current focus of research, and are used in systems like Cassandra, Dynamo and BigTable. Since the arrival of commercially available cloud-computing solutions, such systems can often be deployed over several administrative domains, thus raising the need to ensure trustworthiness of the stored data by design.

The objective of the thesis is the study of scalable and trusted storage and indexing systems, and their different aspects (e.g. scalability, trust management, performance vs. consistency models). The development for the research work is done with the help of the SPLAY framework, which was created in the University of Neuchâtel. Conversely, the code created during the thesis stands for a contribution to SPLAY.

The research on trustworthiness has been done within the Buzzaar project, a joint work with EPFL about distributed aggregation systems. Two complementary systems have been proposed: SPADS & CADA. They ensure anonymity to users and lower the impact of cheating users and servers. The current work involves the developing of a framework to experiment the trade-off between consistency models and availability for storage systems.

The structure of this doctoral thesis can be divided by three chapters. The focus of the first third of the doctoral thesis is to achieve publisher anonymization and rate limitation on clients of a distributed aggregation system.

Many distributed applications, such as collaborative Web mapping, collaborative feedback and ranking, or bug reporting systems, rely on the aggregation of privacy-sensitive information gathered from human users. This information is typically aggregated at servers and later used as the basis for some collaborative service. Expecting that clients trust that the user-centric information will not be used for malevolent purposes is not realistic in a fully distributed setting where nodes are not under the control of a single administrative domain. Moreover, 1 most of the time the origin of the data is of small importance when computing the aggregation onto which these services are based. Trust problems can be evinced by ensuring that the identity of the user is dropped before the data can actually be used, a process called publisher anonymization. Such a property shall be guaranteed even if a set of servers is colluding to spy on some user. This also requires that malevolent users cannot harm the service by sending any number of items without being traceable due to publisher anonymization. Rate limitation and decoupled authentication are the two mechanisms that ensure that these cheating users have a limited impact on the system. SPADS is a system that interfaces to any DHT and supports the three objectives of publisher anonymization, rate limitation and decoupled authentication. A prototype on a cluster was deployed to assess its performance and small footprint.

The second part of the doctoral thesis continues to focus on the security aspects of distributed aggregation systems, taking care of other features that complement the work done on the first part.

For this work, we consider the aggregation of distributions as the base component of a large-scale application operating on a distributed network that spans multiple administrative domains that do not trust each other. The applications are sensitive to biases in the distribution aggregation: the results can only be trusted if inserted values were not altered nor forged, and if nodes collecting the insertions do not arbitrarily modify the aggregation results. In order to increase the level of trust that can be granted to applications, there must be a disincentive for servers to bias the aggregation results. Auditing allows detecting potential misbehavior with respect to the aggregation operation and observed state of the system. The CADA auditing mechanisms let aggregation servers collaboratively and periodically audit one another based on probabilistic tests over server-local state, allowing suspecting malevolent behaviors and supporting lightweight accountability. CADA differs from the existing work on accountability in that it leverages the nature of the operation being performed by the node rather than a general and application-oblivious model of the computation. The effectiveness of CADA is conveyed by an experimental evaluation that studies its ability to detect malevolent behaviors using lightweight auditing oracles.

The last and current chapter of the thesis involves the creation of a flexible distributed storage platform on top of SPLAY, that implements a diverse range state-of-the-art techniques on distribution, replication, routing and indexing on distributed storage systems, and permits the researcher to combine, analyze and compare them through simulation. This will be a base for the research on novel techniques for distributed storage systems, and will give tools to quickly prototype and benchmark the aforementioned novel techniques.

Distributed DNA Alignment, a Stream-Based Approach

Beat Wolf, École d'ingénieurs et d'architectes de Fribourg
Beat.Wolf@hefr.ch

1 Introduction

The field of bioinformatics research becomes becomes more and more popular. Thanks to recent advances in techniques of DNA sequencing, a growing number of genetic data is digitized and analyzed. The collected data is used for diverse purposes, such as diagnostics of genetic diseases, or the mapping of the evolutionary tree of different species. The increasing interest in this technology allowed to decrease the time spent on sequencing a single human genome from about the 13 years achieved by the Human Genome project [5], to as low as 6 hours today. This speed increase has been achieved in only 10 years, which is rarely seen in any technical domain. As a comparison, it is notably faster than the Moore's law, to which computer science is subject to. However this performance improvement created severe processing power problems, which could only partially be resolved by algorithmic advances.

2 DNA sequencing

To understand the problem that needs to be solved, we have to understand the kind of data that is generated by the DNA sequencing and how it is processed. The DNA consists in several chromosomes, which are long sequences of nucleotide bases. The four nucleotide bases are labeled A, C, T and G. The human DNA consists of over 3 billion of nucleotide base pairs. The bases encode the genes of an human organism, and the order of those bases is the information searched by the DNA sequencing process. This process cannot be achieved, for a given human DNA, in one shot. Instead, the DNA is read in many small pieces of 20 to few thousand bases long. These sequences are then aligned against a reference sequence, to find the most likely place each sequence was in the original DNA. It has to be noted that the sequences are rarely identical to the reference sequence and there are multiple regions in the reference sequence which are identical making the alignment process a complex task. Eventually it is the differences in the sequenced DNA compared to the reference sequence which are of interest because they can indicate a disease causing mutation.

3 Distributed alignment

To overcome the increasing gap between the DNA sequencing speed and the speed of the alignment process, new computing approaches need to be explored. As the millions of sequences that need to be aligned are independent tasks, a possible approach is to realize the alignment process on a massively parallel computing infrastructure. Many different distributed architectures have been used to solve this problem, from a simple single multicore machine, to the use of graphics cards (GPUs) through networks of thousands of processing units (Grid). Recently the concept of cloud computing has became a very popular trend. While it removes the burden of individual laboratories to maintain a powerful computing infrastructure, it also raises new questions and problems. One of the major problem arising with the different cloud based solutions for DNA sequence alignment, is the data transfer time from the local network to the cloud, and back again. Tools like Cloudburst [2] or Crossbow [6] need that the data to be analyzed is entirely transferred to the cloud prior to start the computation. This leads to a time overhead of up to several hours before any calculation can be started. In two recent publications, the ETHZ ([3], [4]), has investigated the possibility of using a stream processing approach to overcome the problem of the data transfer time overhead. With this approach, calculations can be started as soon as the first data reaches the computing infrastructure, the data is processed sequence by sequence as and when they arrive.

4 Streaming approach validation

The publication of the ETHZ was based on technologies, that were not necessarily meant to be used in a cloud or streaming based environment. Data conversions had to be made between several incompatible applications which is not good for the overall performance. To validate the concept of streamed alignment, the GensearchNGS aligner, which was created during the master thesis "Analysis and visualization of DNA sequences using cloud computing" [1] and in collaboration with Phenosystems SA, was ported to use a local area network grid through the use of the RMI technology. To test the implementation, a Core 2 Duo laptop clocked at 2.5 GHz was connected to a Phenom X6 1090T clocked at 3.2GHz over a 1Gb/s switch. The laptop has 2 CPU cores and the Phenom X6 has 6 cores. A small dataset of 50k sequences was aligned against the chromosome 1 which is 247 Mega bases long.

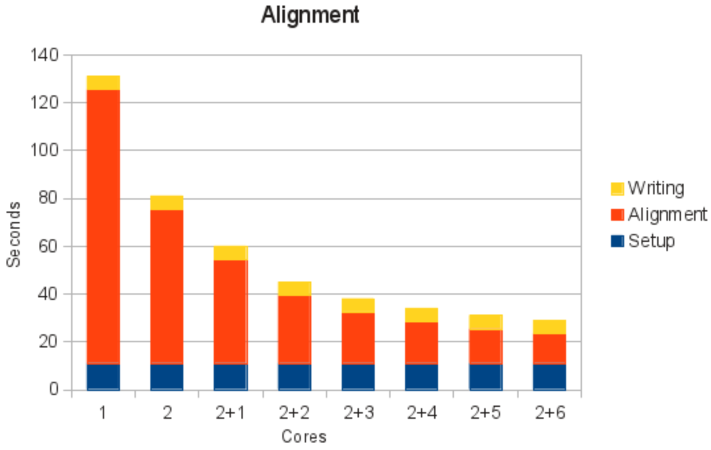


Figure 1: Time needed to align 50k sequences

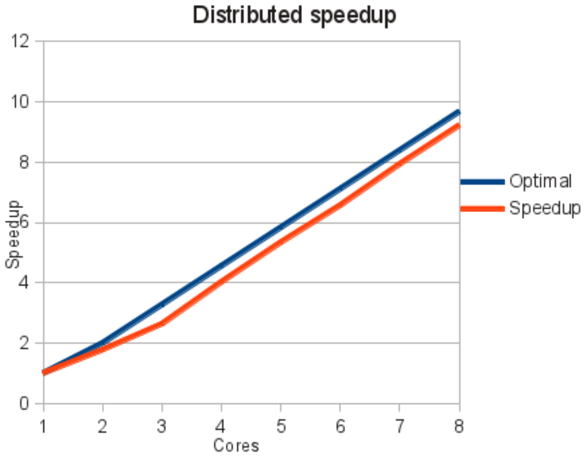


Figure 2: Adjusted speedup over two computers

The results are promising. Even using a technology like RMI that might not be best adapted for the job, a linear speedup was achieved (Figure 2). The main computer for the benchmark was the laptop, the second computer joined as soon as 3 threads were used. The speedup values were adjusted to the power of the processors.

5 Future work

Now that the concept has been validated, more flexible solutions will be explored. Streaming frameworks like DSPE [7], will be looked at because they remove the need to handle implementation complexity of a streaming algorithm and allow the deployment on more diverse architectures. Also looked at will be an implementation of the algorithm using the POP [8] technology, which will enable nearly effortless distributed calculations. Using those technologies will require some improvements. DSPE was developed to run on multicore and GPU accelerated systems. A project is currently being done that combines POP-C++ and DSPE, allowing DSPE applications to run on a POP-C++ powered grid. Further work will need to be done to bring this environment to the cloud.

References

[1] B. Wolf, Analysis and visualization of DNA sequences using cloud computing, 2011

[2] M. C. Schatz, "CloudBurst: Highly Sensitive Read Mapping with MapReduce," Bioinformatics, vol. 25, no. 11, 2009.

[3] R. Kienzler, A. Ranganathan, and N. Tatbul, "Large-scale DNA Sequence Analysis in the Cloud: A Stream-based Approach", 2011.

[4] R. Kienzler, R. Bruggmann, and A. Ranganathan, "Stream As You Go: The Case for Incremental Data Access and Processing in the Cloud", 2012.

[5] Human genome project
`http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml`

[6] "Crossbow
" `http://bowtie-bio.sourceforge.net/crossbow/`

[7] Domain specific language for parallel real-time stream processing
`http://www.systemdesigner.ch/?page_id=31`

[8] POP-C++ and POP-Java
`http://gridgroup.hefr.ch/popc/doku.php/main_page`

Routing, Coverage and Connectivity Issues in UAV ad-hoc Networks

Zhongliang Zhao, Universität Bern
zhao@iam.unibe.ch

Micro Unmanned Aerial Vehicles (UAVs) can be equipped with a variety of sensors to provide aerial communication backbones in many application scenarios. The use of cognitive UAV swarms in applications for distribution of robots and mobile sensors is an appreciable benefit for search-and-rescue and surveillance missions. In this context, the cooperation and coordination between UAVs play an important role in terms of packet transmission and swarm formation control. In this abstract, research issues about routing, coverage and connectivity in UAV ad-hoc networks will be covered.

Advances in electronics and software are allowing the rapid development of small unmanned aerial vehicles (UAVs), capable of performing autonomous coordinated actions. Developments in the area of lithium polymer batteries and carbon fiber-reinforce plastic materials let UAVs become an aerial platform, which can be equipped with a variety of sensors such as cameras. Furthermore, it is also possible to mount communication modules on the UAV platform to let the UAVs work as communication relays to build a wireless aerial backbone network. Recent developments of autonomous unmanned aerial vehicles and wireless sensor networks allow automated approaches to surveillance with minimal human intervention. A feasible solution is to deploy a set of UAVs and each one mounted with a communication module like a wireless mesh node, so as to build a wireless backbone over which various entities on the ground such as rescue teams, relief agencies, first responders, etc. can communicate with each other.

A system of aircrafts would provide mobile ad-hoc networks connecting ground devices with flying UAVs, as well as the inter-connection between different UAVs. One possible approach to achieve this is to maintain a fully connected network of UAVs at all time, so that a given UAV can talk with any other UAV using multi-hop ad-hoc routing. However, oftentimes there are not enough UAVs to establish a continuous path between two points on the ground and this is a huge problem for solutions that require a fully connected UAV mesh network. The notion of a continuous path between end-points is rather simple when the relay nodes are stationary. When the nodes are capable of moving, especially in UAV ad-hoc networks that include nodes moving in a highly mobile way, it becomes extremely difficult to maintain continuous connectivity.

Therefore, a decentralized agent-based motion planning approach is usually applied for UAV controlling. Compared with centralized approaches, autonomous agents are more robust against wireless link failures which might happen due to poor coverage and reliability of cellular technologies in higher altitudes. Since UAVs have to fly with a certain formation to keep connectivity with each other, topology control plays a crucial role.

Due to the high mobility in of UAV ad hoc networks, traditional routing protocols for mobile ad hoc networks (MANETs) may not work properly within such a highly mobile environment. We propose to apply opportunistic routing for packet transmission within such an environment. Simulation results of typical opportunistic routing protocols are proven to be better than traditional MANETs routing protocols [2]. Topology control (coverage and connectivity management) is another essential problem of UAV networks. Three existing approaches with several advantages and drawbacks from distributed agent-based formation control are: Boids Flocking, Potential Field and Virtual Springs. Based on a performed analysis, we propose a topology control solution for UAV ad hoc networks, based on both the received signal strength indicator (RSSI) and GPS location of the flying nodes. Our proposal modifies the Potential

Field approach by considering RSSI values and GPS location data. Rejective forces will be used for collision avoidance. The rejective force will increase for decreasing distance and for increasing RSSI between two UAVs. Attractive forces will be used for main connectivity. The attractive the force will increase for increasing distance and for decreasing RSSI between two UAVs. The GPS data will be taken as backup for lacking channel information or lost connectivity. A weighting of RSSI and distance will be derived accordingly in order to calculate the rejective and attractive forces [1].

References

[1] Z. Zhao and T. Braun, "Topology Control and Mobility Strategy for UAV Ad-hoc Networks: A Survey," Joint ERCIM eMobility and MobiSensor Workshop, Santorini, Greece, June 8, 2012.

[2] Z. Zhao, B. Mosler and T. Braun, "Performance Evaluation of Opportunistic Routing Protocols: A Framework-based Approach using OMNet++", Submitted to the 7th Latin America Networking Conference 2012, Medellin, Colombia, October 4-5, 2012.