Measurement-Based Handover Method for Communication-Based Train Control Systems

Mei Ya Chan[1], Sami Baroudi[1], Joseph Siu[2], Jörg Liebeherr[1]

[1] Department of Electrical and Computer Engineering, University of Toronto
[2] Thales Canada Transportation Solutions, Toronto

***Abstract*—In Communication-Based Train Control (CBTC) systems, trains communicate with wayside access points (APs) placed along a track using 802.11 or similar wireless technologies. A critical component of the communication system is the handover algorithm used by a train to select an AP as it moves along a track. As a safety-critical system, the number of handovers should be kept small and selected APs must satisfy given QoS requirements. We present a novel measurement-based handover algorithm for CBTC systems that exploits the predictable motion paths of trains. Using empirical measurements from a rail communication system, we show that our algorithm keeps the number of required handovers small, while avoiding ping-pong handovers, i.e., situations where a train bounces back and forth between the same access points.**

***Index Terms*—CBTC System, Handover Algorithm, Received Signal Strength**

I. Introduction

In railway transit systems, such as light rail and subway systems, trains are scheduled on a track in a pipelined manner to ensure safe distances between consecutive trains. Safety distances are enforced by assigning each train a movement authority which is the permission to move through a designated section or block of track at a specified speed. Movement authorities are computed and updated based on the current location and speed of each train. For this, it is pivotal that trains communicate their state information to a central control station and receive, in return, a movement authority within a stringent time frame. Modern train control systems that incorporate radio communication technology to exchange this and other essential information are called Communication-Based Train Control (CBTC) systems [7], [9]. Since train control messages are time-critical, communications in CBTC systems are subject to stringent latency and availability requirements to assure railway safety, e.g., as specified in IEEE 1474 [1].

The main components of a CBTC system are a wired communication network that connects a central control station to a number of wayside APs, which have been deployed along the track, and a train equipped with wireless radios to associate with a wayside AP. Figure 1 presents an illustration of a CBTC system. For redundancy, trains are typically equipped with two radios, one in the front and one in the rear of the train, that each select an AP independently. Throughout this paper, we only consider that each train has one radio mounted at the front of the train.

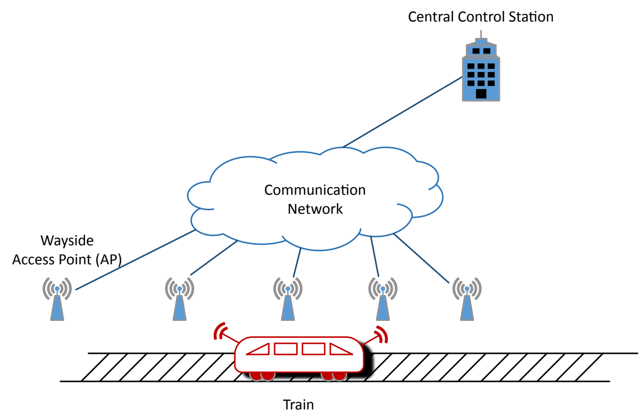


Fig. 1: Architecture of a CBTC system.

As the train moves along its track, the association with its current AP degrades and it has to select a new AP. The process of transferring the association from one AP to another is called handoff or handover. A method for determining when to perform a handover and for selecting a new AP is called a handover method. Since CBTC systems are time-critical, it is vital that a handover decision results in the selection of an AP that satisfies minimal Quality-of-Service (QoS) requirements. Since each handover bears the risk of a failure, handover decisions for CBTC systems seek to keep the number of handovers small. In order to achieve seamless communication with high QoS between a train and the central control station, the main challenge is to perform handovers to the right APs at the right time.

In addition to minimizing the number of handovers, a handover method also seeks to minimize durations of service interruption, defined as time durations when a selected AP experiences a failure or degraded service due to hardware failures, weak signals, or strong interference. Finally, a handover method for CBTC systems should eliminate ping-pong handovers, which refer to situations where a train bounces back and forth between the same APs. Satisfying the opposing design goals of simultaneously decreasing the number of handovers while also reacting quickly to a degraded service, all in the presence of train mobility, make the design of a handover method for CBTC systems challenging.

In this paper, we present a measurement-based handover method for CBTC systems. The novelty of our method is that it exploits the largely deterministic motion path of trains on the same track. Since trains operating on the same track experience a similar QoS when passing the same location, measuring the QoS of a track and making the measurements available to subsequent trains can improve handover decisions. We propose the creation of a database which records the measurements taken by trains and creates a statistical model for the experienced QoS. The statistical model is used to create a plan for handover locations, which we refer to as *handoff map*. By continuously searching for deviations of current measurements from the statistical model, a train can react quickly to unexpected events, such as failures or sudden QoS degradation, and recompute the handoff map. In this way, the handover method becomes resilient to network failures and sudden degradations of service.

This paper is organized as follows. In Section II we present our method for exploiting predictable motion paths of trains. In Section III we develop a statistical model for the link quality on a train track. In Section IV we discuss the proposed handover method. In Section V we present an evaluation using measurements of a CBTC system. We present brief conclusions in Section VI.

II. Exploiting Predictable Motion Path

Since they involve only a single type of access technology, handovers in a CBTC system between trains and wayside access points are called horizontal handovers. The literature on horizontal handovers for mobile wireless networks largely ignores location information. There exists a limited body of work on location-aware handover schemes [4], [6], [11], one even targeting a CBTC deployment, however, solutions that exploit situations where the motion path of mobile nodes is predictable do not exist.

Our scheme exploits that a train travels on a fixed track resulting in a deterministic motion path. In Figure 2 we illustrate a time-space diagram, where arrows indicate the movement of trains as a function of time. The three trains indicated in the figure maintain similar speeds and stop at the same locations. The figure suggests that trains operating on the same track encounter a similar wireless environment. We exploit this characteristic by developing a handover method that uses knowledge of the position and coverage of deployed APs. The knowledge base is created by previous trains that have moved along the same track. Using measurements of the link quality of the deployed APs, we can develop a statistical model that shows the range of positions where an AP provides an acceptable QoS level, referred to as the *coverage* of the AP. Once the coverage of APs is estimated, a handoff map can be created by linking APs that have overlapping coverage, with the condition that any point is covered by at least one AP. (To increase resilience to failures, one can add the requirement that each point on the track must be covered by two or more APs.) In this way, trains, before they start on their route, can determine the APs to hand off to and the exact locations where

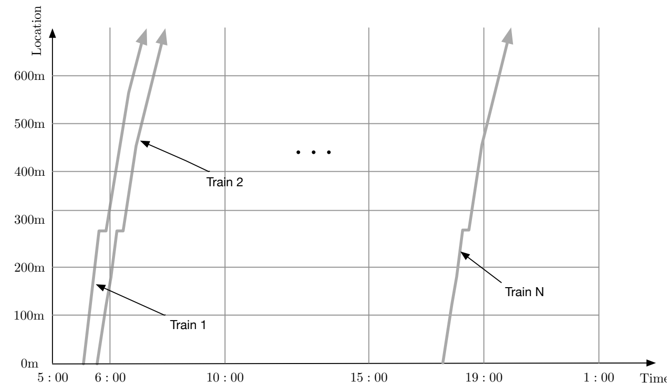


Fig. 2: Time-space diagram.

the handovers will be performed. In the following, we will refer to the traversal of a whole track by a train as a 'trip'.

In the event of a hardware failure, a hardware degradation, or a physical obstruction between an AP and a train, measurements may deviate significantly from the statistical model. We refer to time periods that show statistically significant deviations between the current measurements and the statistical model as *unpredictable events*. The operation of a measurement-based handover scheme trains must continuously perform tests to detect the presence of unpredictable events. If an unpredictable event is detected for an AP that is included in the current handoff map, the train computes a new handoff map that excludes the affected AP.

III. Link Quality Model

For our study, we use the Received Signal Strength (RSS) as the metric to indicate the achievable QoS between a train and an access points. This is supported by [12], which evaluates the effectiveness of Received Signal Strength (RSS), Signal to Interference plus Noise Ratio (SINR), Packet Delivery Ratio (PDR) and Bit Error Rate (BER), and found that the use of RSS is an effective metric. The RSS, which measures the power of a received signal, is easily measured and generally available from the radio platform.

We consider the development of a statistical model for a track with deployed wayside APs. The spatial model of the location on a train track can be simplified to one dimension. We assume that the track has a length of D, and we measure locations using the distance from the start of the track. We denote the set of all wayside APs deployed along the track by $\Phi = \{\phi_1, \phi_2, ...\}$. To account for spatial variations, we view RSS measurements of a specific AP ϕ as a spatial series $\{X_d^\phi : 0 \leq d \leq D\}$ where d denotes the location where the measurement is taken. The difference between the time series $X_{d_1}^\phi$ and $X_{d_2}^\phi$ for two locations $d_1 \neq d_2$ captures changes of path loss, shadowing and multipath fading. Spatial variations may occur on a much smaller scale than what can be supported by the hardware. For example, the state change when entering a tunnel due to multipath fading is in the order of a few wavelengths, corresponding to less than a meter at 2.4 GHz. However, hardware limits the sampling rate, yielding a distance between RSS samples that is in the order of tens of meters. Therefore, we discretize the track in coarser-grain segments, and maintain a spatial series for each segment. In this paper, we assume that segments have a constant length of n meters, resulting in $\lceil \frac{D}{n} \rceil$ segments. In our experiments, we assume $n = 10$ m. For each segment j, we maintain a single spatial series X_j^ϕ, which stores all measurements taken at locations $(j-1)n < d \leq jn$. The selection of optimal segment sizes is addressed in [3].

When a train traverses a track, it collects the RSS of APs at a given rate r. When a train completes a trip, the train has collected a spatial sequence $\{X_d^\phi : 0 \leq d \leq D\}$ for each access point ϕ that has been measured. Each such sequence is a sample path that is mapped to the per-segment spatial series X_j^ϕ by converting d to a segment j. The data collection is repeated until the statistics of X_j^ϕ for each segment j can be unbiasedly and consistently estimated.

We interpret the spatial series of segment j, X_j^ϕ, as samples of an unknown random variable. The samples of a series yield the empirical distribution of X_j^ϕ. We say that an AP ϕ covers a location d if $Prob(X_j^\phi > C^{\text{thresh}}) \geq p_c$, where C^{thresh} is the coverage threshold and p_c is a required confidence level.

The threshold C^{thresh} is an estimate of the minimum acceptable RSS level to maintain a good connection by achieving a target data rate. The threshold value C^{thresh} can be adjusted based on the noise floor and interference at different locations, e.g., open space, tunnel, or train station. For example, 802.11b requires a minimum RSS level of -85 dBm and a minimum SINR of 4dB for a target rate of 1 Mbps [8]. Suppose the noise plus interference level at a certain location is -90 dBm, then C^{thresh} can be set to $\max(-85, -90) + 4 = -81$ dBm.

When an AP ϕ becomes unavailable, suffers a degradation, or experiences another unpredictable event, X_j^ϕ is no longer an unbiased estimator. We detect unpredictable events by comparing measured samples with the empirical distribution of X_j^ϕ for a segment. If the empirical distribution makes recent values of the measured quality unlikely, we can conclude that an unpredictable event has occurred. A deviation is registered if a recent measurement x satisfies $Prob(X_j^\phi \leq x) \leq p_{u0}$, for a small probability p_{u0}. Note that we are only interested in negative deviations. Measurements where the quality is higher than expected are ignored. Since a single data point may be an outlier and not truly indicate an abnormal condition, we trigger an unpredictable event only if multiple consecutive measurements result in negative deviations. If it is valid to assume that events $\{X_j^\phi \leq x\}$ are independent, the probability of N consecutive deviations is $p_{u_0}^N$. Denoting the cumulative distribution function of X_j^ϕ by $F_{X_j^\phi}$, applying the quantile function yields $x \leq F_{X_j^\phi}^{-1}(p_{u_0})$. If N consecutive measurements satisfy the inequality, we conclude that an unpredictable event has occurred. For our evaluations we use $N = 3$.

IV. Measurement-Based Handover Method

In this section we describe our methodology for measurement-based handovers for a CBTC system. We consider a wireless wayside network where APs are deployed along the track. We require that trains are able to keep track of time and their location. Localization can be achieved by means of transponder tags along the track, a tachometer, or GPS. The methodology has two phases:

- *Measurement phase:* The purpose of the measurement phase is to collect sufficient data to justify the statistical model of the link quality of APs.
- *Operation phase:* This phase outlines the actions during regular operations of a train as part of the CBTC system.

We envision that trains in the operations phase continue to take measurements to update and enhance the statistical model.

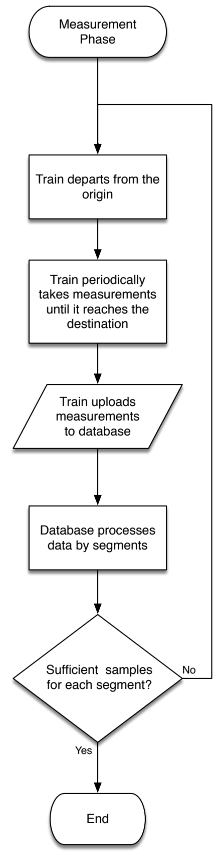
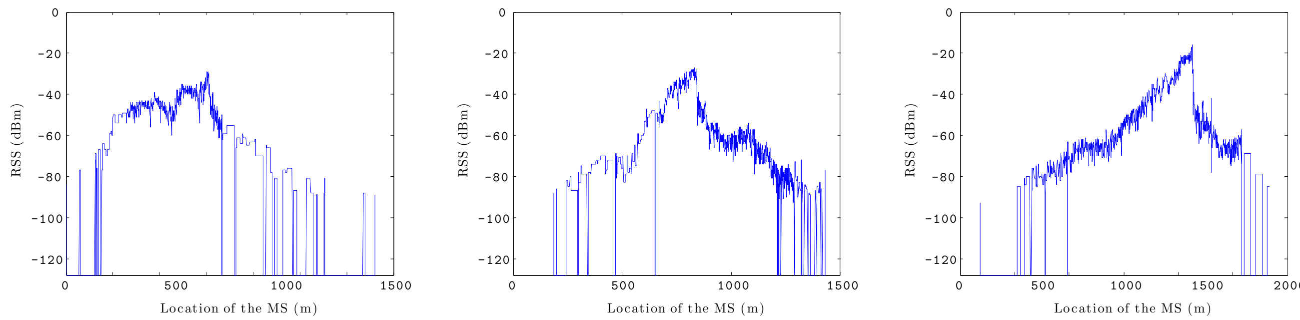


Fig. 3: Measurement phase.

A. Measurement Phase

The measurement phase consists of repeated measurements of APs from trains that traverse a track. In Figure 3 we present a flow diagram of the measurement phase. A train departs at the beginning of a track, the *origin*, and traverses the track. The on-board radio equipment of a train periodically performs a scanning process where a radio searches for nearby APs and measures the QoS of the APs. Each recorded data point consists of a tuple (time, location, *AP*, *RSS*), where *RSS* is the measured QoS for access point *AP*, at a given time and location. Note that the location is one-dimensional, and consists of the distance from the origin. At the end of the track, the measurement data is uploaded to a database, which organizes the data into segments, and generates the QoS model for each AP. The scan results are accumulated and processed



(a) AP 16. (b) AP 19. (c) AP 22.

Fig. 4: Measurements of APs on a track.

by segments. The measurements of the measurement phase are repeated until sufficient samples for each segment have been collected, so that unpredictable events can be reliably detected. Our measurements are based on radio units that generate 4 to 5 scans per second, where each scan can make RSS measurements of up to 8 APs. If a train is moving at 50 km/h, there are 28 to 36 data points per 100 m. This means that measurements of about 30 trips are needed to collect 1000 data points for each segment.

In Figure 4, we show the results of RSS measurements of a train (referred to as MS in the figure) of selected APs from a CBTC system. The measurements are taken on a train line with a length of approximately 2.4 km, which is equipped with 32 wayside access points. Drops to -128 dBm indicate incidents of missing data. The figures give some indication how the RSS values depend on the location at which they are measured.

As discussed in earlier sections, the measurements of the RSS at a segment are viewed as a random variable whose distribution is characterized through experimental measurements. For the operational phase, we extract from the distribution two threshold values. The first value, used to determine the coverage of an AP and denoted by C_j, is the 10^{th} percentile value of the empirical distribution X_j^{ϕ}. The second value, called the unpredictable RSS value and denoted by U_j, is set to the 0.001^{th} percentile value of X_j^{ϕ}.

Even though CBTC systems require only about 50 kbps data rate for each train as reported in [10], additional requirements, e.g., for real-time voice communication, result in a minimum data rate of about 1 Mbps. The minimum RSS sensitivity levels and minimum SNR values required depend on the technology of the radios. For example, commodity 802.11b devices requires a minimum sensitivity level of -85dBm and a minimum SNR of 4dB for 1Mbps, 7dB for 2Mbps, 11dB for 5.5Mbps, and 16dB for 11Mbps [8].

B. Operation Phase

The process of the operation phase is summarized in Figure 5. The operation phase is divided into two parts. Before starting on a trip, a train downloads a copy of the database and computes the coverage profile of all APs on the track. This is used to create a handoff map with a minimum number of handovers, subject to meeting redundancy constraints. While moving on the track, the train follows the handoff map and performs handovers at the locations determined by the map. During the trip, a train continuously takes measurements and compares them with the data from the database. If there is a statistical deviation between measurements and the database, the presence of an unpredictable is concluded. Before taking action on the event, the train confirms its presence with the central control station. If the event is confirmed and if the event occurs for an AP that is included in the current handoff map, a new handoff map is created.

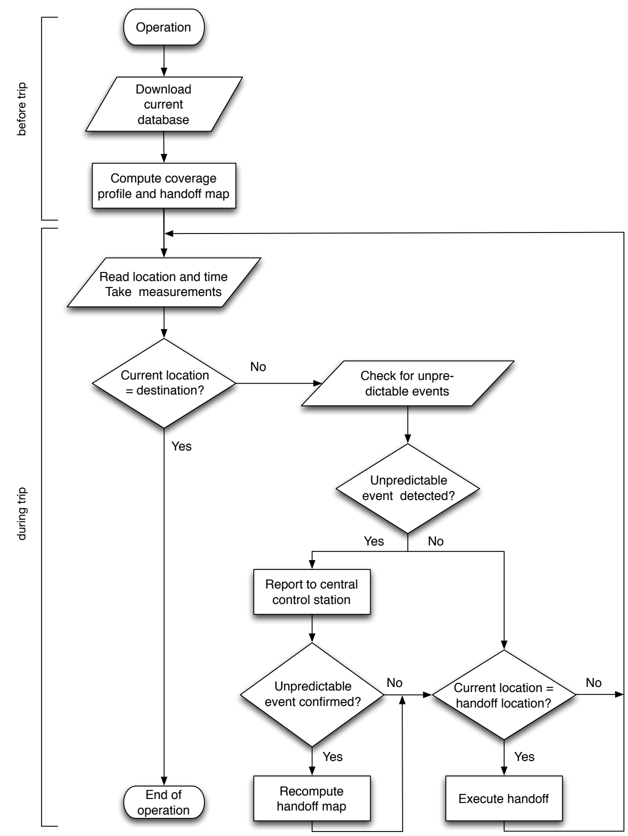


Fig. 5: Operation phase.

In the following, we discuss the coverage profile, handoff map, and detection of unpredictable events in more detail.

Coverage Profile: The coverage profile of an AP is generated by comparing the percentile C_j of the AP for a segment j to the threshold value C^{thresh}. Note that, by comparing C_j with C^{thresh}, we are in fact working with the Signal-to-Noise Ratio (SNR), which is a better estimation of channel quality than the absolute RSS level. With our parameter selection, if $C_j > C^{\text{thresh}}$, there is a 90% likelihood that the measured RSS exceeds C^{thresh}. In this case, we mark the segment as 'available' for the measured AP. Otherwise, we mark it as 'unavailable'. The maximum consecutive sequence of available segments makes up the coverage of an AP. As seen in Figure 4, the measured RSS of an AP increases as the train approaches the AP, and decreases after it has passed the AP. Therefore, applying a cut-off threshold generally results in one consecutive sequence of available segments. The estimated coverage of all APs makes up the coverage profile.

Handoff Map: Once the coverage areas of APs have been determined, the creation of the handoff map appears straightforward. However, there are additional considerations that must be taken into consideration. To account for the handover delay, i.e., the time required to form an association to a new AP, we require the coverage areas of two APs to overlap by at least 100 m. This corresponds to about 5 seconds at the maximum speed of 70 km/h. A safety margin of 5 seconds is sufficient, as measurement studies have shown that horizontal handovers in IEEE 802.11 consume less than 0.5 seconds [5], whereas the execution of a vertical handover between IEEE 802.11 and GPRS takes an average of 2.37 seconds [2]. Then, we find a map that yields the minimum number of handovers by finding the shortest path from an AP at the current location (generally, the origin) to an AP at the destination location. There will generally be a large number of solutions for the shortest path. These solutions are ranked by the length of the overlapping coverage. To balance the network load, the handoff map creation can be further refined by having consecutive trains select different handoff maps. Even if the trains know the APs of a handoff map, executing handovers too early or too late may result in handover failures. If the overlapping coverage area of two consecutive APs in the handoff map has multiple segments, the segment where the handover is selected is the segment where the expected RSS of the next AP exceeds that of the previous AP.

Unpredictable Events: If a train detects $N = 3$ consecutive RSS measurements in segment j that are all below U_j, an unpredictable event is triggered. With U_j set to the 0.001^{th} percentile, observing three consecutive values below the threshold corresponds to a probability less than 10^{-9}. Note that taking three measurements at a rate of 4 to 5 samples per second corresponds to 0.6 to 0.75 seconds. If an unpredictable event is detected for the currently associated AP or the next AP in the handoff map, the handoff map is recomputed by removing the affected AP from the coverage profile. Unpredictable events at other APs do not require a change of the handoff map.

V. Evaluation of the Handover Method

We conduct an evaluation of the handover method using a measurement trace of a light rail system that has a length of 2.4 km with 32 deployed APs. The trace consists of the RSS values of the eight strongest APs in range of the train, which are measured every 0.2 to 0.25 seconds. The location of the train is computed from timestamps using a separate data set that logs the location of the train at certain times. We assume that the trains move at a constant speed between the log entries. The measurement traces are used both for the database as well as the RSS values measured by a train. We refer [3] for additional details on the methodology. Figure 4 shows the data for three of the APs obtained from the measurement trace.

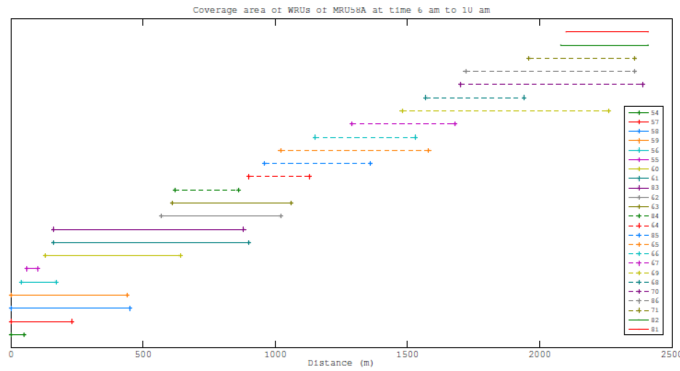


Fig. 6: Coverage profile of APs.

In Figure 6, we show the coverage areas of 24 APs, where we set the threshold value to $C^{\text{thresh}} = -70$ dBm. While the coverage of individual APs varies widely, most APs serve multiple segments. From the figure, we see that the deployment of APs is sufficiently dense to allow numerous handoff maps. Using the method described in the previous section, the shortest path computation satisfying a minimum overlap of coverage of APs, results in the following handoff map (including the computed handover locations):

Access Point:	AP58	AP83	AP62	AP64
Location:	Start	223 m	662 m	1005 m

Access Point:	AP65	AP69	AP82
Location:	1109m	1559 m	2165 m

We now compare the performance of the measurement-based handover methods with two other methods, one that is entirely based on RSS and one that is entirely based on location. The evaluation is conducted in a simulated wireless environment, where the length of the track, the movement path of a train, and the number of access points are identical to those of the measurement trace mentioned above. RSS value readings are equally separated by 0.2 seconds. The list of APs in each scan includes all APs in the system. An RSS measurement is generated for each AP using path loss and shadowing, where the path loss exponent is set to 2, and the shadow standard deviation is set to 4 dB. All APs are assumed to have identical radio hardware. The receiver sensitivity of each AP is assumed to be -84 dBm, that is, an AP is undetectable if the RSS falls below the given sensitivity. The evaluation compares the number of handovers and the length of the handover intervals, while keeping track of time periods with service interruptions (time periods when the associated

TABLE I: Performance comparison.

Performance Metric:	RSS -based	Location -based	Our method
Number of handovers	7	20	4
Average handover Interval (m)	290.3	118.0	465.3
Number of service interruptions	0	0	0
Number of ping-pong handovers	0	0	0
Number of low QoS incidents	116	0	24
Max. low QoS duration (s)	0.8	0	0.4

AP is undetectable) and low QoS (time periods when the RSS of the associated AP falls below -70 dBm)

The RSS-based method starts by selecting the AP with the highest RSS and preserves an association until the RSS drops below $C^{\text{thresh}} = -70$ dBm. If no AP is found with a better RSS, the current association is maintained. The location-based method always makes an association to the nearest AP ahead. Both methods are reactive in the sense that they rely entirely on current measurements.

The results are summarized in Table I. We observe that our measurement-based methods performs the least number of handovers, resulting in longer intervals between handovers. All methods avoid service interruption and ping-pong handovers. The advantage of the frequent handovers with the location-based method is that it can avoid periods of low QoS. Both the RSS-based and our method show several periods when the associated AP falls below the desired threshold value. On the other hand, the periods of low QoS are limited to time periods of 0.8 seconds for the RSS-based method and 0.4 seconds for our method. The advantage of our method is that it keeps the number of handovers small, while maintaining the time periods of a low QoS level to two measurement intervals.

VI. Conclusions

We have presented a handover method for a CBTC system that exploits that trains have predictable motion paths. We propose to use measured data to determine the statistics of the coverage of wayside access point, and perform advance planning using a handoff map that determines the locations of handovers between access points. We have compared the measurement-based handover method with two reactive handover schemes. Our measurement-based method showed a reduction in the required number of handovers. We believe that taking into consideration the predictable motion paths of trains can help to reduce the cost of a CBTC system by reducing requirements for deployed wayside access points.

Acknowledgements

This work has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by Thales Canada Transportation Solutions.

References

[1] IEEE 1474.1-2004. Standard for Communications-Based Train Control (CBTC) performance and functional requirements, 2004.

[2] M. Bernaschi, F. Cacace, and G. Iannello. Vertical handoff performance in heterogeneous networks. In *Proc. International Conference on Parallel Processing Workshops*, pages 100–107, August 2004.

[3] M. Y. Chan. Enhancing Communication Availability with Hybrid Networks. Master's thesis, University of Toronto, Canada, 2016.

[4] H. Jiang, H. Zhao, and B. Zhao. A novel handover scheme in wireless LAN in CBTC system. In *Proc. IEEE International Conference on Service Operations, Logistics, and Informatics*, pages 473–477, July 2011.

[5] A. Mishra, M. Shin, and W. Arbaugh. An empirical analysis of the IEEE 802.11 MAC layer handoff process. *ACM SIGCOMM Computer Communication Review*, 33(2):93–102, 2003.

[6] J. Montavont and T. Noel. IEEE 802.11 handovers assisted by GPS information. In *Proc. IEEE Conference on Wireless and Mobile Computing, Networking and Communications*, pages 166–172, June 2006.

[7] S. Morar. Evolution of communication based train control worldwide. In *Proc. Professional Development Course on Railway Signalling and Control Systems (RSCS)*, pages 218–226. IET, 2012.

[8] H. Q. Nguyen, F. Baccelli, and D. Kofman. A stochastic geometry analysis of dense IEEE 802.11 networks. In *Proc. IEEE Infocom*, pages 1199–1207, May 2007.

[9] R. D. Pascoe and T. N. Eichorn. What is communication-based train control? *IEEE Vehicular Technology Magazine*, 4(4):16–21, 2009.

[10] Siemens Transportation Systems. High availability train to wayside communication system for metro applications. Siemens, 2008.

[11] C.-C. Tseng, K.-H. Chi, M. D. Hsieh Deng, and H.-H. Chang. Location-based fast handoff for 802.11 networks. *IEEE Communications Letters*, 9(4):304–306, 2005.

[12] A. Vlavianos, L. K. Law, I. Broustis, S. Krishnamurthy, and M. Faloutsos. Assessing link quality in IEEE 802.11 wireless networks: Which is the right metric? In *Proc. IEEE Personal, Indoor and Mobile Radio Communications*, pages 1–6, September 2008.