MACH-T: A Behavior-based Mobile Node Trust Evaluation Algorithm

Karen H. Thurston
Computer and Office Studies Department
Long Beach City College
kthurston@ieee.org

Daniel Conte de Leon
Center for Secure and Dependable Systems
University of Idaho
dcontedeleon@ieee.org

Abstract

Resiliency and availability in community and public service networks may be economically enhanced by building new ad hoc networks of private mobile devices and joining these to public service networks at specific trusted points. Resiliency in such ad hoc networks relies on the afforded increased availability but also on security which is in turn built on trust. In this article, we describe MACH-T, a novel behavior-based algorithm for mobile ad hoc network node trust building. MACH-T uses historical mobile node geographic location behavior to incrementally calculate node trust values based on the concepts of node capability, commitment, and consistency. We describe experiments and results from evaluating MACH-T using real GPS traces from the Microsoft Research GeoLife and University of Rome Tor Vergata Roma Taxi datasets. Our results show that MACH-T builds a reliable trust value and corresponding confidence value based on learned patterns of time spent in qualifying geographic locations.

Keywords: algorithms, trust, networks, spatio-temporal data mining, location history

1. Introduction

Mobile devices have become ubiquitous. Wireless capabilities for device-to-device communication are varied and currently available in most mobile devices. Applications such as *FireChat* and *BluetoothChat* have seen widespread use in areas where cellular service has been interrupted or is non-existent, demonstrating the viability of proximal group communication using mobile phones not relying on cellular infrastructure (Thurston, et al., 2019). Mobile phones are able to communicate using Bluetooth at distances of about 100 meters. With sufficient density of devices and predictable presence of devices in specific geographic locations, groups of devices can form mobile ad hoc networks known as MANets. For example, research into delay tolerant networks comprised of mobile devices not using cellular infrastructure has shown message delivery rates of 80% to 92% within 72 hours within the geographic area of a college campus (Liu, 2016).

Applications such as *FireChat* and *BluetoothChat* assume a binary trust determination based on human mobile device users or owners knowing each other and/or assuming trust a priori. Apple corporation's *Find My* Network or Amazon corporation's *Sidewalk* network use large numbers of smartphones and smart-home devices, respectively, to provide services not relying solely on cellular networks (Mims, 2021). These two examples also require known identity authentication as a condition for joining the network.

1.1. The problem

Authentication-based trust determinations are well-suited for network applications that involve human intervention such as chat applications, human-to-human messaging applications, or proprietary and closed systems, such as *FireChat*, *BluetoothChat*, *Find My* and *Sidewalk*. However, requiring a priori trust determination is not possible in many other interesting applications, particularly in MANets. For example, (1) ad hoc IoT device message forwarding, (2) managed messaging to, from, within, or across ad hoc networks with thousands or more devices or nodes, (3) emergency response message forwarding and delivery, (4) community networks and community-service message forwarding and delivery. Hence, trust in ad hoc networks, and for these types of applications cannot be based on binary authentication and must be built based on other parameters.

This paper proposes to answer the question, "Can movement and communication history of mobile nodes provide a reliable measure of trust?"

1.2. The Contributions

This paper provides three contributions:

1) Demonstrates a novel method for measuring mobile device trust based on movement and communication behavior.

2) Presents results of implementing an algorithm called MACH-T (Movement And Communication History for Trust) for measuring mobile device trust.

URI: https://hdl.handle.net/10125/103259
978-0-9981331-6-4

3) Presents the results of two experiments for evaluating the MACH-T approach for the cases of personal and taxi-mounted mobile device GPS (Global Positioning Satellite) traces (chronological record of device locations given in longitude and latitude coordinates).

The experiments were not an effort to determine if the unexpected or outlier behavior is malicious, only that it is undesired and therefore not trustworthy. Trust is context dependent and can be defined by the trustor in any way desired. The same trustee may be trusted by one trustor and not another.

The MACH-T algorithm provides an approach and method for calculating trust and a confidence value for adjusting the trust value. MACH-T translates the trust attributes of capability, commitment, and consistency into the following measurable attributes:

Capability: Data availability and longevity;

Commitment: Repeated visits within a geographic location for a minimum duration;

Consistency: Repeated visits to a small number of locations.

We envision a future system where builders and operators of MANets would be able to:

1) Select certain geographic areas of interest for building a MANet, and

2) Select mobile nodes within those geographic areas based on calculated trust values using MACH-T.

This article describes our approach, algorithm, experiments, and results toward enabling point 2) above by using geographical behavior of mobile devices to build trust values. Geography is a critical criterion when mobile nodes act as alternate infrastructure in place of built systems such as the cellular system. Hence, location behavior patterns are an important basis for a MANet operator to determine which nodes should be allowed to join the MANet.

We used real mobile device GPS traces from the Microsoft GeoLife (Microsoft Research, 2012) and University of Rome Tor Vergata Roma Taxi datasets (Bracciale, et al., 2017) to analyze and discover predictable geographical behaviors. These behaviors were used as the basis for constructing a generic trust evaluation algorithm, MACH-T. MACH-T was then evaluated against the GeoLife and Roma Taxi datasets showing promising results.

1.3. Organization of this Paper

The rest of this paper includes the following sections: Section 2 describes related work and knowledge gaps, Section 3 describes the GeoLife and Roma Taxi datasets, Section 4 describes our approach and methods, Section 5 describes our experiments and results, Section 6 presents a discussion, Section 7 presents the conclusion, Section 8 provides acknowledgements followed by Section 9 References.

2. Related Work and Knowledge Gaps

The subject of trust has been addressed in many disciplines including sociology, economics, philosophy, psychology, organizational management, and autonomic computing in industrial and system engineering.

In MANets, the five characteristics of trust are that it is dynamic, subjective, not necessarily transitive, asymmetric (need not be reciprocal), and context dependent. These characteristics define a relationship between cooperating nodes in a MANet and like trust in human relationships, are complex and not easily measured. As it relates to security, trust can be considered a prerequisite or a result of security as in "trustworthy" (Cho, et al., 2010).

In exploring various trust management schemes for MANets, Cho, et al. found *"no work clearly addresses what should be measured to evaluate network trust."* For individual node trust metrics, Cho, et al. proposed future research to include measuring both social reputation and quality of service. This research does not explore social reputation but does address quality of service by measuring factors contributing to quality as defined by the operator of a MANet. Section 4: Approach and Methods provides details about these factors.

Establishing trust in mobile devices has been the subject of study from the time mobile devices first began to proliferate, in the early part of the twenty first century. Trust is the basis for security and privacy of communications.

This research demonstrates an approach for calculating some important attributes of trust of a network node based on its location behaviors prior to allowing it to join a MANet. One similar implementation of this concept is the online social network NextDoor.com which requires users receive a postcard with a confirmation code at their residence address to ensure they live in the neighborhood where they claim to live.

In Security Metrics: Replacing Fear, Uncertainty and Doubt, the advice is "Trust is good, control is better" (Jaquith, 2007). To control system variables such as which nodes are allowed to join the MANet, trust must be measurable. As former US president Ronald Reagan once quoted a Russian proverb, "Trust, but verify." In other words, observe or measure whether trust is warranted. This research provides a method to measure mobile node trust.

In large cities, the density of mobile devices carried by humans or attached to other moving carriers has made

device-to-device communications a viable option for various schemes including IoT sensor data collection and distribution of software updates to devices (Shah, et al., 2003). Mobile phone data can provide insights into population density to inform which locations are candidates for such services (Deville, et al., 2014). This research, by measuring the movements of mobile nodes, can provide a MANet operator with the information required to select specific nodes that frequent a particular geographic location, increasing the viability of a MANet in that location.

In 5G cellular technology rollouts a study proposed methods to offload cellular network traffic to device-to-device networks using femtocell technology to augment the macro cellular infrastructure in a Heterogeneous Cellular Network (HCN) (Lu, et al., 2018). Trust, however, in this study was only dependent on authorization, key authentication, prior social interactions, or device-to-device performance in the already formed network, not on historical device location behaviors or other measurable attributes.

Various proposed architectures for MANets include a layer to oversee the management of the MANet. The term "Trust Overlay Network" originally proposed metrics for determining reputation of nodes in an established MANet (Zhou, et al., 2006). The MACH-T method described in this paper can be classified as a "Trust Overlay Network" with the reputation established by empirical measurement of location behavior prior to a node joining the MANet.

Eagle and Pentland (2009) in the behavioral sciences have shown predictability in human behaviors and they use the term *eigenbehaviors.* These can predict, with a high degree of certainty, where someone will be in the future, based on past behavior. This research demonstrates an approach, although not based on the same eigenbehavior method, to assign trust values to mobile devices carried by human operators based on past behavior.

Although wireless communications rely on unencrypted broadcast protocols, encrypted cellular network communications are not without risk. Reports of the United States Federal Bureau of Investigation (FBI), and municipal law enforcement agencies setting up cell-site simulators, also known as IMSI-catchers (international mobile subscriber identity) or stingrays to capture communications of suspected criminals has also captured communications content of cellular calls between non-suspected users (Ney, et al., 2017). The public using cellular infrastructure may not be aware of this specific risk to privacy although widespread popular media coverage of security breaches in general has raised awareness of security and privacy concerns with mobile phone data, but mobile phone usage has only increased year over year, not decreased. The *Signal* messaging app which claims to provide more secure mobile phone communications can use mobile phone Bluetooth protocol when phones are near to each other, but only provides authentication by registering users with mobile phone telephone numbers from cellular providers and uses multi-factor authentication when the user registers for the service. There is no consideration given to location history of the phone, at least it is not given as a criterion for registering for the service. Conversely, this research demonstrates how mobile nodes could form networks of trusted nodes apart from the cellular network.

A United States patent application proposes using geolocation and timing for issuing a one-time password for authentication purposes but does not address the issue of network node trust over time (Agarwal , 2019). This was one of only two technologies found that used geolocation in any way, but unlike this research, did not use geolocation traces to establish measurable trust.

Node trust computations are simpler in static networks because node behavior is predictable after enough observations, but mobile node trust computations are hard when the location is constantly changing (Govindan and Mohapatra, 2011). This research shows that mobile node trust can be calculated when nodes are moving in the case of nodes carried by human operators or mounted in vehicles. This research found that there are measurable group averages for how many times mobile nodes move and are stationary during a 24-hour period. This behavior can also be said to be predictable for a given group.

A proposed "general theory of trust" is based on "human expectations and mental models of trust without relying on false metaphors and analogies with the physical world." Trustworthiness should factor in computational correctness and a behavior trust primitive (Gligor and Wing, 2011). This research demonstrates that the physical world is crucial because mobile nodes must be close together for a MANet to be viable. Knowing which nodes are repeatedly present for a minimum amount of time in a specific location is crucial to inform a MANet operator.

Another study assumed nodes were trustworthy prior to joining a network after a "bootstrapping" period in research to evaluate the trustworthiness of nodes based on artificially induced node interactions, including requests for assistance between nodes (Saied, et al., 2013). This research does not assess the quality of intranode communication, only that nodes are present and able to communicate as evidenced by GPS trace history.

In the book *Modeling Trust Context in Networks*, identity and authentication are given to be important for security (Adali, 2013). Adali acknowledges the work of another author and says identities on networks result in

"disembodiment" and if identity is not "tied to … physical presence" it is a barrier to establishing trust (Nissenbaum, 2004). This research does tie physical presence to trust and demonstrates a way to measure trust.

A NIST (National Institute of Standards and Technology) publication described a proof-of-concept implementation of the Trusted Platform Module (TPM) configured to know the current location of the hardware to enforce geolocation restrictions for purposes of restricting computing devices outside of national borders, for example, but not to track the geographic locations of a moving hardware device over time (NIST, 2015). The Trusted Platform Module is a hardware device installed on computers that cannot be compromised by malicious software and verifies that the operating system (the platform) on a computer has not been compromised. This research does not demonstrate hardware implementations to measure trust, but the NIST publication supports the idea that geolocation can be used as a reliable method for establishing trust.

Other researchers have investigated trust for mobile ad hoc networks. However, we found no academic reports describing an approach in which trust is built based on actual historical geographical behavior plus context-sensitive requirements established by MANet operators such as requirements for presence in a geographic location or trust confidence thresholds.

A survey and classification of trust computation models identifies several gaps in trust research. One subtopic area, within the topic of trust, for which a marked lack of published research results appears to exist, is when there are several distinct trust metrics contributing to one overall trust value. Only four papers focused in this area and only considered social and distributed, peer-to-peer types of trust building (Guo, et al., 2017).

By contrast, the novel algorithm MACH-T and the experimental results we describe in this article contribute new knowledge in the subtopic area of *multi-attribute trust formation* identified by Guo, et al. Our research also has the potential for developing Trust-as-a-Service (TaaS), another approach needing further research as suggested by Guo, et al.

Within the Guo, et al. trust model classification, one model, which they call "Class 5: QoS + social / distributed / static weighted sum / event + time-driven / multi-trust" would fit the closest to the MACH-T model presented in this article. Our model differs from Guo et al. Class 5, in that it does not have a social attribute component since these attributes tend to assume direct involvement by humans as an essential portion of the trust computation. In addition, our trust formation model is intended to be automatic and centralized rather than manual and distributed.

Also, Guo et al. evaluated trust computation models as one of: trust composition, trust propagation, trust aggregation, trust update, and trust formation. Our approach falls closer to their classifications as trust formation and trust update. Furthermore, MACH-T is built to support the concepts of "capability", "commitment", and "consistency" that result in trust such as in human relationships when one person or group considers another person or group to be capable, committed, and consistent in their behavior (Hacker, 2014). Although our approach does not use human or device recommenders to directly contribute to trust ratings, the empirical data we analyzed considers the same factors. This is because the mobile nodes we analyzed have human operators and we assume devices reflect the behaviors of those humans even though in our case do not have humans directly involved as recommenders.

Table 1 provides a mapping of the trust classification design dimensions from Guo, et al. to the human behavior dimension from Hacker showing full coverage between the computing and human behavior domains. This table provides the basis for the MACH-T formula described in Section 4: Approach and Methods.

Table 1. Computing trust classification design mapped to human behavior trust

Trust Dimension in Computing Domain (Guo et al., 2017)	Trust Dimension in Human Behavior Domain (Hacker, 2014)
Quality of Service	Capable: Location collected frequently over time by communicating with GPS satellites/tracking devices
Centralized	Consistent: Centralizing provides comparison to others/self to determine consistency in behavior
Static Weighted Sum	Consistent: Repeated conformance to population average or ideal geographic location behaviors
Event + Time-driven	Committed: Repeated visits to locations over time
Multi-trust	Capable: Many dimensions contribute to overall trust

3. Datasets

The GeoLife dataset from Microsoft Research (Microsoft, 2012) is anonymized and publicly available. Microsoft researchers collected GPS traces from 182 subjects mostly between April 2007 and August 2012. The average subject's device data spanned 6.2 months (standard deviation of 1 year, 2.4 months). Microsoft's statistics for the dataset are 17,621 trajectories with a

total distance of more than 1.29 million kilometers and 50,176 hours. The sampling rates vary from 1 to 5 seconds and between 5-10 meters per data point. The data was mostly gathered in and around Beijing, China, with some traces in the United States and Europe. Each subject's traces contained latitude, longitude, altitude (not used by this research), and date.

The RomaTaxi dataset from Bracciale, et al. (Bracciale, 2014) is also anonymized and publicly available. The researchers collected GPS traces from 291 taxi-mounted GPS tracker Android devices during a six-month period from October 2013 through April, 2014, in Rome, Italy. The first four days of data from February, 2014 are available for public download. The average subject's data spanned 2.46 days (standard deviation of 1.091 days). The sampling rates averaged 15.164 seconds (standard deviation of 1.104 seconds). The data was mostly gathered in central Rome, Italy. The researchers stated their focus was on an 8km x 8km area (or 64 km^2) in the center of Rome.

The average geographic area traveled in this analysis of the dataset was 63km^2 for subjects with MACH-T(A) values greater than zero with a 20-minute stay requirement. Some subjects traveled far outside the focus area as in the case of trusted subjects in the 10 minute stay requirement analysis with an average area covered of 98.2km^2. Untrusted subjects in both analyses traveled even farther, a maximum of 1,166km^2 in the 10 minute stay analysis. Researchers who collected the data stated they collected traces every 7 seconds, but the dataset we download contained traces every 15 seconds.

The University of Idaho Office of Research Assurances through its Institutional Review Board (IRB) evaluated the details of the described datasets and the proposed research and determined that these activities did not meet the definition of Human Subjects Research and do not require IRB oversight.

4. Approach and Methods

Repeatable time in qualifying locations is a key factor used in the MACH-T method for determining the trust attributes of capability, commitment, and consistency. In this section we describe our trust formation and confidence formulae.

4.1. Approach

The first step in our evaluation methodology was to analyze the GPS traces of both datasets to determine average behavior for time in unique locations for each dataset's population. We anticipated that most GeoLife subjects would spend at least 20 or 60 minutes in only a handful of unique locations over time, such as home, work, school, places of worship, shopping, and recreation. We calculated the total hours and number of times a subject visited the same location for at least the minimum time duration. Because taxis move more frequently and faster than pedestrians, we analyzed their behavior using both 10 and 20 minutes as the required minimum duration in a location. Average behavior provided for some of the "Consistency" aspect of trust. A node behaving in a manner consistent with the general population of nodes is desirable.

We assumed that most mobile devices either carried by human owners or mounted on vehicles would not be actively mobile for more than a portion of a 24-hour period and would incur periods of being stationery. For example, for human carried mobile devices, even in more congested areas with long commutes such as the Los Angeles basin in California, USA, where average commute times are 31 minutes each way, or about 1 hour per day, commute time accounts for only 4.2% of a 24-hour day (California, 2022).

Hence, rather than focusing on when mobile devices were moving, we focused our analysis on the times when devices were stationary within a given location. We defined a location as a square of approximately 469 x 469 meters square (this parameter, zoom level 16, is adjustable). We designated a minimum elapsed time that a device must be in a given location (this parameter is also adjustable) in order for that location to qualify as a valid location.

We did not restrict analysis to any given date range within all dates in the data which ranged from April 2007 to the end of July 2012 for GeoLife subjects, and during February 2014 for Roma Taxi subjects. A MANet operator may want to restrict analysis to a specific number of days in the immediate past.

4.2. Data Processing

We conducted our analysis by designing, coding, and executing a C++ program to read each of the trace files for each subject to record one or more locations for each subject for any continuous time in the location of at least 10, 20, or 60 minutes (adjustable). For any two trace records where the time interval between any two GPS trace records was longer than 10 minutes (adjustable), we did not add that time to the accumulated time in one location, assuming the GPS tracking application was either disabled or lost contact with the GPS satellite. We added intervals longer than 10 minutes (adjustable) to the total of all trace intervals for a subject as an indicator of the total time span during a day from the first trace to the last. Assuming a MANet may rely on devices being able to respond to messages within a brief time, we chose 10 minutes, but this value could be smaller or larger depending on the needs of the MANet operator. The 10-minute requirement also increased the confidence value of the calculated trust value.

4.3. Tiling the Geo World Map

The C++ program converted latitude and longitude input values in the GPS trace files to x,y tile coordinates using the OpenStreetMap.org (OpenStreetMap, 2022) conversion algorithm. This algorithm assigns each location on the earth to a grid location (tile) through a method known as spherical pseudo-Mercator projection providing non-overlapping relatively square locations approximately 469 meters on each side at the latitude of Beijing, China at a zoom level of 16. Zoom level 0 represents the whole earth. Level 16 divides the earth into 4,294,967,296 tiles.

The tiles become smaller at more northern latitudes due to the curvature of Earth and the spherical pseudo-Mercator projection does not adjust for this factor. Because most of our data points were near Beijing, China and Rome, Italy, which are at similar latitudes, we used the tile size of 469 meters on each side as a constant size. Future versions of our C++ program could adjust the tile size according to latitude. The Open Street Maps documentation does not provide standard measurements for on the ground distances between tiles at various zoom levels. To calculate this distance as 469 meters, we used the GeoFabrik Tile Calculator web resource (Geofabrik, 2020) and the GPSprune (GPSprune, 2020) Windows application.

4.4. Calculated Fields

Table 2 shows the data format for the summary fields we calculated for each subject's qualifying locations. We calculated summary data for each subject to establish population averages and standard deviations. The *Hour* and *DOW* fields are for future use.

Table 2. Qualifying location record for a subject

Attribute Name	Attribute Description
xTile	X tile coordinate
yTile	Y tile coordinate
Hour	Beginning hour of day
DOW	Day of the week
Freq	Number of times in a location for the minimum qualifying time
Hours Duration	Total number of hours in a location
Trace Count	Number of trace records for a location
First Date	First/oldest date in a location
Last Date	Last/most recent date in a location

4.5. The Unadjusted Trust Value

Here we describe in detail each term of the formula shown in Fig. 1 for the MACH-T_U trust calculation.

1. The ratio of qualified hours (QH: number of hours greater than the minimum required time in the same location) to qualified days (QD: number of days with at least one qualified location) is an indication of a node's capability to be present. No qualifying days is zero trust.

2.The ratio of qualified locations (QL: number of locations with more than the minimum required time in the same location) to qualified days (QD) is an indication of a node's capability to be present. If a node has no qualifying locations, the value is zero.

3. The ratio of qualified days (QD) to total days (TD: number of GPS trace days) is an indication of a node's commitment to action over time. If TD is zero, the node has no GPS trace data, the trust value is zero.

4. The ratio of qualified location area in km^2 (QL) to the perimeter area in km^2 (QL Perim) is an indication of a node's consistency in behavior. Visiting qualifying locations within a small total perimeter area indicates high suitability for a MANet; a node's presence is highly predictable within a constrained perimeter.

5. The ratio of qualified locations (QL) to total locations (TL) is an indication of a node's consistency in behavior. If a node's qualifying locations are a large % of all the node's locations, it is an indication of a node's consistency in behavior and high suitability for a MANet; the node's availability is highly predictable.

6. The ratio of qualified hours (QH) to total hours (TH) is an indication of a node's commitment to be available for MANet operation.

7. Term denominators: All the denominators for each term are the average plus one standard deviation of the same ratio in the numerator. If the node's numerator value is greater than the average + 1σ of all nodes in the population, this factor will increase the MACH-T_U value more than if the value is less than the average + 1σ.

8. Weights: The formula in Fig. 2 has six terms and six weighting factors. For our experiments we equally weighted these six weights at 16.66%. A MANet operator might adjust these weights for each factor depending on the needs of a planned ad hoc network.

$$\text{MACH-}T_U =$$
$$(1)\left[W_1 * \frac{\frac{QH}{QD}}{\left(\frac{\overline{QH}}{\overline{QD}}\right)+1\sigma}\right] +$$
$$(2)\left[W_2 * \frac{\frac{QL}{QD}}{\left(\frac{\overline{QL}}{\overline{QD}}\right)+1\sigma}\right] +$$
$$(3)\left[W_3 * \frac{\frac{QD}{TD}}{\left(\frac{\overline{QD}}{\overline{TD}}\right)+1\sigma}\right] +$$
$$(4)\left[W_4 \frac{\frac{QL\,km^2}{QL\,perimter\,km^2}}{\left(\frac{\overline{QL\,km^2}}{QL\,permimeter\;km^2}\right)+1\sigma}\right] +$$
$$(5)\left[W_5 * \frac{\frac{QL}{TL}}{\left(\frac{\overline{QL}}{\overline{TL}}\right)+1\sigma}\right] +$$
$$(6)\left[W_6 * \frac{\frac{QH}{TH}}{\left(\frac{\overline{QH}}{\overline{TH}}\right)+1\sigma}\right]$$

Figure 1. The MACH-T_U trust algorithm formula

4.6. Confidence and Adjusted Trust

We also calculated a confidence value for each node that depends on the density and volume of the available GPS trace data. We then calculated an adjusted trust value, MACH-T_A, to account for the confidence level from the trace data using the formula in Figure 2.

For the experiment described in this article, we chose 10 as the number of desired traces per hour and 30 as the number of desired days in a date range. However, these parameters may be adjusted as desired depending on the objectives and risk tolerance of the MANet operator.

We describe in detail each term of the formula shown in Figure 2 for the CONFIDENCE value calculation.

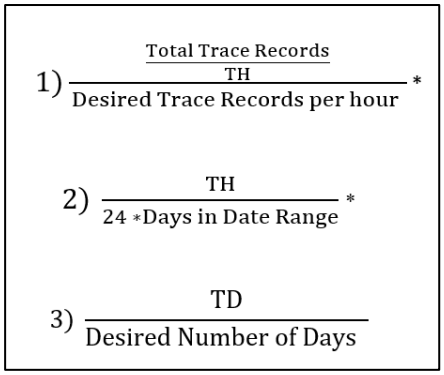

Figure 2. Confidence formula

1) This term determines the ability of a node to communicate at a desired rate when the node is communicating. It does not reflect a node's ability to be available on any given day. Maximum confidence would be for subjects with trace records at a desired rate such as every 10 seconds for all days and hours within the start and end date range for which they had trace records. Depending on the desired confidence level, a MANet operator may choose a maximum trace interval between individual trace records before assuming a trace segment has ended. In our experiment we chose to restart the time and trace counter if a trace interval was longer than 600 seconds (10 minutes). Because our zoom level defined areas of approximately 469 meters on a side, for 10 minutes it would be possible for a subject to leave the area and return later to record another trace which could be interpreted as being in the same place, but the subject would need to stay close to the area to accomplish this within the 10-minute time limit.

2) This term is the total possible hours (TH) during the date range of trace records represents the maximum possible trace data. Traces representing a high percentage of all possible hours during a date range (24 hours per day in the range) will have a higher confidence and indicate a node's ability to be present in qualified locations within a 24-hour period.

3) This term is the total days (TD) in the trace data as a percentage of the desired days reflects a node's ability to be present within a date range but does not reflect frequency of communication during an average day.

Using the confidence value, we then calculated the adjusted trust value, MACH-T_A as:

$$MACH\text{-}T_A = MACH\text{-}T_U * CONFIDENCE$$

5. Results and Analysis

5.1. Analysis of GeoLife GPS Traces

Although each subject created GPS trace data on their own schedule and at days and times of their own choosing, results nevertheless confirmed observable and repeatable behavioral patterns for most of the subjects for visits of a minimum time duration to the same geographic location visited by the subject over time.

Table 3 shows the number of trusted subjects (out of 182 subjects) where the minimum stay requirement in a location was 20 minutes. Trusted subjects are defined as having a MACH-T_A value greater than zero. The trusted nodes' first six qualifying locations accounted for 91.9% of all qualifying locations, indicating a high probably that a node would be found in one of only six locations when stationary for at least 20 minutes.

Table 3. Distribution of locations: GeoLife

Number of Qualifying Locations	20 Minute Minimum Stay (129 trusted subjects)
1	54.7%
2	72.8%
3	80.7%
4	85.7%
5	89.4%
6	91.9%

The following tables show results of analyses of the GeoLife GPS trace files for 182 subjects. Table 4 lists the totals, averages, minimums, maximums, standard deviations, and modes for the summary data for all 182 subjects. Table 5 shows the coefficient values calculated for using in the MACH-T trust formula. The trust coefficients shown in Table 5 were calculated using the values from Table 4. Table 6 shows the MACH-T values for the top 10 trusted subjects or devices. The table shows the values for the unadjusted trust, MACH-T_U, the Confidence, and the adjust trust, MACH-T_A for each subject in descending MACH-T_A value order.

Table 4. Observed population statistics: GeoLife

	Total Days (TD)	Total Hours (TH)	Total Locations (TL)	Qualified Locations (QL)	Total Qualified Hours (QH)	Total Qualifie Days (QD)
Total	10634	9288	485154	836	706.3	908
Average	58.4	51.0	2665.7	4.6	3.9	5.0
Minimum	1	0.03	1	0	0	0
Maximum	1245	1286.5	99104	69	51.1	65
Standard Deviation (σ)	129.1	125.7	8765.3	8.3	6.8	8.9
Mode	6	7.7	58	0	0	0

Table 5. Trust formula coefficients: GeoLife

Calculated Trust Formula Coefficients	QH/ QD	QL /QD	QD /TD	QL Km² / QL perimeter km²	QL/ TL	QH/ TH
Averages: $\overline{QH/QD}$, $\overline{QL/QD}$, $\overline{QD/TD}$, $\overline{QL/TL}$, $\overline{QH/TH}$ **and** $\left(\frac{QL\ km^2}{QL\ permimeter\ km^2}\right)$	.554	.767	.116	.239	0.006	0.096
Averages + 1σ	1.095	1.453	0.28	0.646	0.024	0.207

Table 6. Top 10 of 129 trusted GeoLife subjects

Subject	MACH-T (U)	Confidence	MACH-T (A)
151	2.1545	0.0664	0.1430
160	1.0918	0.1251	0.1366
143	1.5775	0.0713	0.1125
137	1.3178	0.0818	0.1078
41	0.3759	0.2448	0.0920
48	0.8703	0.0977	0.0850
77	1.3075	0.0433	0.0567
141	1.2547	0.0432	0.0542
123	1.0718	0.0464	0.0497
25	0.3897	0.1210	0.0472

5.2. Analysis of Roma Taxi GPS Traces

The Roma Taxi dataset spans four days in February 2014, concentrated in mostly a small geographic area less than 100km^2. When requiring 20-minute minimum stays, trusted subjects in the Roma Taxi dataset spent 98.6% percent of qualified hours in the first six qualified locations as shown in Table 7.

Table 7. Distribution of locations: Roma taxis

Number of Qualifying Locations	20 Minute Minimum Stay (259 trusted subjects)
1	54.4%
2	76.2%
3	88.0%
4	94.1%
5	97.2%
6	98.6%

The following tables show results of analyses of the Roma Taxi GPS trace files for 291 subjects. Table 8 lists the totals, averages, minimums, maximums, standard deviations, and modes for the summary data for all 182 subjects. Table 9 shows the coefficient values calculated for using in the MACH-T trust formula. The trust coefficients shown in Table 9 were calculated using the values from Table 8.

Table 9 contains the coefficient values arbitrarily set as ideal coefficients. The trust coefficients are not based on population averages as they were with the GeoLife experiment. The Roma Taxi experiments instead used arbitrary ideal values to use in the MACH-T_U formula.

Table 10 shows the MACH-T values for the top 10 trusted subjects or devices. The table shows the values for the unadjusted trust, MACH-T_U, the Confidence, and the adjusted trust, MACH-T_A for each subject in descending MACH-T value order. Using arbitrary ideal coefficients resulted in trust values slightly greater than 1.0 in six cases.

Finally, Figure 3 shows the geographic area covered by the trusted Roma Taxi subjects.

6. Discussion

6.1. Scope of Analysis

Our analysis did not consider any geographic location as more desirable than any other location, but MANet operators could impose an additional capability factor for this attribute to find mobile devices in specific locations. Selection of a desired geographic location could be one of the dynamic weighted sum variables used at run time to calculate trust.

Table 8. Observed population statistics: Roma taxis

	Total Days (TD)	Total Hours (TH)	Total Locations (TL)	Qualified Locations (QL)	Total Qualified Hours (QH)	Total Qualified Days (QD)
Total	716	3110	115688	981	958.9	503.0
Average	2.46	10.7	397.6	3.37	3.3	1.73
Minimum	1.0	0.0	1.0	0.0	0.0	0.0
Maximum	4.0	30.9	1574.0	11.0	14.57	4.0
Standard Deviation (σ)	1.09	6.89	286.98	2.31	2.85	1.09
Mode	3.0	9.0	326.0	3.0	0.0	1.0

Table 9. Ideal trust formula coefficients: Roma taxis

Trust Formula Coefficients	QH/ QD	QL /QD	QD /TD	QL km² / QL perimeter km²	QL/ TL	QH/ TH
Arbitrary Ideal Values →	6.0	2.0	1.0	0.064	0.004	0.5

TABLE 10. Top 10 of 259 trusted Roma taxi subjects

Subject	MACH-T (U)	Confidence	MACH-T (A)
308	3.3084	0.3715	1.2292
70	3.1518	0.3818	1.2033
151	3.5179	0.3325	1.1696
212	3.2424	0.3510	1.1382
141	3.1343	0.3538	1.1090
48	3.0587	0.3335	1.0201
269	2.9562	0.2991	0.8843
356	2.3471	0.3700	0.8684
244	3.0074	0.2877	0.8652
186	2.2206	0.3771	0.8374

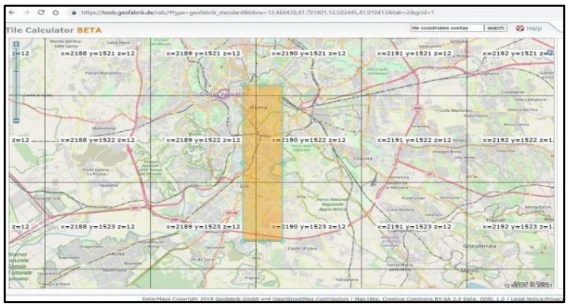

Figure 3. Area of trusted Roma taxi subjects

Using the MACH-T approach, a MANet operator would be able to classify mobile devices according to any given expectation of how frequently and for how long those devices ought to be at a certain set of selected locations. The MANet operator defines the size and coordinates of the locations. The level of confidence on that classification is also defined by the MANet operator. Also note that these are virtual networks, defined by a grouping of mobile devices. A given universe of devices could be used to form many MANets each formed from different geolocation trust criteria, used alone or in combination with other classification criterion, as needed for the network objectives. For example, if the purpose of the network is to forward IoT sensor data in a specific city, forming a MANet with devices of owners currently living in that city would help the network achieve its objective. The key point is the devices exhibit behavior over time that is consistent (the same over time), capable (able to communicate), and committed (for a certain minimum time period).

We reserved the hour and day of week values for future use to simplify the results, but these values would be potential additional terms in the trust formula for networks operating only at certain hours of the day or days of the week. The day of the month and the month of the year could also be specified and contribute to the trust rating.

Additionally, we did not restrict the age of GPS trace data to a range of dates in the recent past. Some GeoLife data was collected as far back as 2007 and we considered that data equally with more current data from 2012. MANet operators may decide to impose a period for collecting behavior data such as for example the most recent past 30 days. Behavior data in the recent past could be weighted more heavily than data in the distant past, but long term consistent desirable behavior should be weighted heavily as it supports all fundamental aspects of trust as our method defines it.

7. Conclusions

Our experimental results show GPS traces provide sufficient data to calculate a trust measure in mobile devices based on geographic movement history, answering our research question, "Can parameters such as movement and communication history of mobile nodes provide a reliable measure of trust?"

In our experiments, and for the subjects we deemed to be trusted, given our input parameters, over 90% of trusted devices spent all their stays of 20 minutes or more in only six locations. The significance of this small number of locations is a positive indicator for the formation of MANets which require persistent presence within a geographic area small enough to allow mobile devices to communicate in the absence of cellular infrastructure.

MACH-T, a novel method for calculating a trust metric, is flexible enough to support diverse types of mobile networks and devices as long as patterns of presence in certain arbitrary locations can be observed from the device population.

Before ad hoc networks can become secure and resilient, trust will need to be measurable and dynamic. Historical and geographical behavior analysis is a promising avenue for providing the basis for device (node) trust formation in mobile ad hoc networks. Other trust criteria such as the requirement to be present in a specific geographic location or locations can be added to the evaluation algorithm as desired by the network operator. The MACH-T algorithm does not include a term for specific location presence, only the criteria to exhibit population typical behavior: visiting a limited number of locations over time for a minimum duration. A MANet operator might want to include only certain geographies for a MANet and would disqualify/not trust nodes not visiting specific geographies.

Additionally, point in time trust values can be dynamically updated as desired. Potential security threats to a MANet built using the MACH-T trust algorithm will be the subject of future research.

8. Acknowledgments

The authors thank Long Beach City College and University of Idaho staff who ensure support services and research infrastructure are available as needed. The authors also thank the HICSS-56 conference committee, mini-track and track chairs, and reviewers for their help improving this paper. We thank Dr. Thurston's dissertation committee: Dr. Lori Baker-Eveleth, Dr. Frederick Sheldon, and Dr. John Shovic for providing valuable feedback on this research. The authors thank Long Beach City College, the State of Idaho, and the National Science Foundation (NSF #1565572) for partially funding this research and its presentation. The opinions expressed in this paper are not necessarily those of Long Beach City College, the State of Idaho, or the National Science Foundation.

9. References

Adali, S. (2013). Modeling Trust Context in Networks. Springer Briefs.

Agarwal, G. (2019). U.S. Patent Appl. No. 15/920,973.

Bartock, et al. (2015). Trusted Geolocation in the Cloud. NISTIR 7904.

Bracciale, et al. (2014). CRAWDAD Dataset Roma/Taxi (v. 2014-07-17), doi: 10.15783/C7QC7M.

California Average Commute Time by County. https://www.indexmundi.com/facts/united-states/ quick-facts/california/average-commute-time

Cho, et al. (2010). A Survey on Trust Management for Mobile Ad Hoc Networks. *IEEE Communications Surveys & Tutorials*, *13*(4), 562-583.

Deville, et al. (2014). Dynamic Population Mapping using Mobile Phone Data. Proceedings of the National Academy of Sciences, 111(45), 15888-5893.

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying Structure in Routine. Behavioral Ecology and Sociobiology, 63(7), 1057-1066.

Garrett, et al. (2021). The Acceptability and Uptake of Smartphone Tracking for COVID-19 in Australia. *Plos One*, *16*(1), e0244827.

GeoFabrik Tile Calculator (2020) https://tools.geofabrik.de

Gligor & Wing (2011). Towards a Theory of Trust in Networks of Humans and Computers. In *International Workshop on Security Protocols* (pp. 223-242). Springer, Berlin.

Govindan & Mohapatra (2011). Trust Computations and Trust Dynamics in Mobile Ad hoc Networks: A survey. *IEEE Comm. Surveys & Tutor*., *14*(2), 279-298.

Liu, Y. (2016). Supporting Large Scale Communication Systems on Infrastructureless Networks Composed of Commodity Mobile Devices: Practicality, Scalability, and Security (Doctoral dissertation).

GPSprune. https://activityworkshop.net/software/gpsprune

Guo, J., Chen, R., & Tsai, J. J. (2017). A Survey of Trust Computation Models for Service Management in Internet of Things Systems. *Computer Communications*, *97*, 1-14.

S. Hacker. Who Do You Trust? Quality Progress 47.8 (2014).

Jaquith, A. (2007). *Security Metrics: Replacing Fear, Uncertainty, and Doubt*. Pearson Education.

Lu, et al. (2018). Trusted Device-to-Device based Heterogeneous Cellular Networks. IEEE Transactions on Vehicular Technology, 67(11), 11219-11233.

Microsoft Research. (2012). GeoLife GPS Trajectories. http://research.microsoft.com/en-us/downloads/ b16d359d-d164-469e-9fd4-daa38f2b2e13/

Mims, C. (2021). Thanks for Powering our Wireless Network ... Wall Street Journal May 08, 2021, ProQuest.

Ney, et al. (2017). SeaGlass: Enabling City-wide IMSI-Catcher Detection. Proceedings on Privacy Enhancing Technologies, 2017(3), 39-56.

Nissenbaum, H. (2004). Will Security Enhance Trust Online, or Supplant it? *Trust and Distrust Within Organizations: Emerging Perspectives, Enduring Questions, Eds. R. Kramer and K.Cook,Russell. Sage Publications*,155-188.

OpenStreetMap.org. https://www.openstreetmap.org

Saied, et al. (2013) Trust Management System Design for the Internet of Things:", Comput. Secur. 39, 351–365.

Shah, et al. (2003). Data Mules: Modeling and Analysis of a Three-tier Architecture for Sparse Sensor Networks. *Ad Hoc Networks*, *1*(2-3), 215-233.

Thurston, K. H., & de Leon, D. C. (2019). MACH-2K Architecture: Building Mobile Device Trust and Utility for Emergency Response Networks. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)* (pp. 152-157). IEEE.

Zhou & Hwang (2006). Trust Overlay Networks for Global Reputation Aggregation in P2P Grid Computing. In *Proc. 20th IEEE Intl' Parallel & Distributed Processing Symposium* (pp. 10+). IEEE.