RESCUE MUNI

Muni Riders' Survey Results

San Francisco, CA

March 19, 1997

Muni Riders' Survey Results

Summary

Perhaps the most common complaint about the San Francisco Municipal Railway (Muni) is that it is unreliable. While it does come very close to covering every area of the city, according to many riders Muni runs late very often, making it difficult if not impossible to depend on Muni for anything time-sensitive.

RESCUE Muni, the San Francisco transit riders' association, designed the Muni Riders' Survey to assess the severity of this problem from the rider's perspective. For fifteen days in early February (2/1 to 2/15), RESCUE Muni members and other volunteers recorded how long they waited for the buses or streetcars they used every day. We then compared these times with the frequencies published on maps and shelters and confirmed our suspicions: Muni riders continue to experience delays at an unacceptable rate.

As recorded by Riders' Survey participants, Municipal Railway riders experienced delays at least 25 percent of the time during the survey. (For 25 percent of all rides taken, the participant waited longer than the frequency advertised by Muni.) This means that a commuter who takes Muni to and from work every weekday can expect to be delayed every other day. (Riders who transfer once a day can expect to be delayed every day.) While this was the system average, many routes were far less reliable. Several routes were late more than 40 percent of the time, and the worst route of all that had a sufficient sample, Route 22-Fillmore, was late a whopping 55 percent of the time.

Other issues particularly affecting commuters were demonstrated by this survey. For example, as we expected, the system was less reliable during rush hour than at other times, running 27% late in the morning and 28% late in the evening. Expresses, also running 28% late, behaved similarly. (Midday commuters had better results, with Muni running only 21% late.) This may help to explain the particular dissatisfaction with Muni among daily commuters in San Francisco. It may also explain in part why Muni ridership has declined substantially in recent years: according to Muni's own statistics, ridership has declined by approximately 11 percent in the last four years.[1]

In Table 1, we have listed overall system performance as well as a selection of routes. Later in this paper, we provide a complete ranking of routes, and we also provide performance comparisons based on criteria such as the time of day and the type of route. Throughout this discussion, we will use letter grades (using the normal teachers' scale) to describe lines' on-time performance; under this system, Muni's on-time performance as a whole rates a C. (22-Fillmore, for example, rates an F.)

[1] SF Municipal Railway, *Short Range Transit Plan 1996-2005*, 1996, p. 3.7.

Table 1: Summary of Riders' Survey findings

route	*% late*	*grade*	*# responses*
System-wide total	25%	C	1365
Worst 5 routes:			
22-Fillmore	55%	F	33
1-California	43%	F	51
19-Polk	42%	F	45
F-Market	39%	D	33
15-Third	34%	D	50
Best 5 routes:			
5-Fulton	16%	B	44
54-Felton	10%	A	21
6-Parnassus	9%	A	45
KLM-Muni Metro	7%	A	58
27-Bryant	5%	A	22

This Survey does not attempt to determine the root causes of these delays, though our participants were generous with their comments and suggestions for improvement. After describing the findings in detail, we will discuss some possible reasons for delay brought up by our participants, but we believe that it is fundamentally Muni's responsibility to find and solve problems that impact on-time performance. In particular, Muni should be able to provide *consistent* performance across all lines.

Methodology

We conducted the Muni Riders' Survey for fifteen days, from February 1 to February 15. Participants, mostly but not all members and supporters of RESCUE Muni, recorded for each bus or streetcar they took, the following information:

- route
- location
- date
- time they arrived at the stop
- time the vehicle arrived

Some participants also recorded the vehicle number, which we are not using for this study but which may prove useful in the future.

After data were submitted by mail or our World Wide Web site, we entered this information into our database using Microsoft Excel. For each data point, we calculated waiting time and compared it with the frequency advertised on Muni shelters and in the

Official San Francisco Street & Transit Map, 1996 edition.[2] If waiting time exceeded the advertised frequency, that vehicle was late from the rider's perspective, even if it met some internal Muni schedule. We used this Boolean variable to calculate the percentage late for each route and for other criteria such as time of day and type of vehicle.[3]

This is an admittedly generous standard, because riders can generally be expected to arrive at the bus stop at random intervals. (Riders who arrive after others have waited for some time may wait less than the advertised frequency and therefore not report a late bus as late.) We chose this standard because it was a simple one to administer and because it reports the minimum number of buses and streetcars that are late. Our hypothesis, which seems to have been proven by the data, was that Muni is so unreliable that it will still show a poor on-time performance record despite the generous standard.

For each ride taken, we also measured waiting time and a "normalized waiting time," waiting time expressed as a percentage of the published frequency. These are also included in the analysis of specific lines and of the system as a whole. (Please see the Technical Appendix for a detailed description of these metrics.)

In total, 97 people participated in the survey, generating 1365 data points. The most widely reported line was the N-Judah streetcar, with 120 responses, followed by the L Taraval line with 59 and the KLM underground Muni Metro line with 58; we will use these particularly well-ridden lines, as well as some others, as the basis for more detailed analysis of performance below.

Note: The data in this survey are quite comparable to previous studies conducted by Muni and RESCUE. The major difference is that this study is done from the rider's perspective: in our opinion, it is much more important to understand how riders are affected by Muni failures than it is to assess Muni's performance against internal benchmarks. RESCUE Muni conducted a brief study of several lines in November 1996 in which observers tracked the arrival times of six routes for one hour each. In that survey, 46 percent of vehicles were found to be late in comparison to the published schedule; when we imported these data into our database, we found that the typical rider would at that time have been delayed 18 percent of the time. Our conclusion from this comparison is that Muni has not improved at all since last fall; in fact, it may have worsened a bit.4

---

[2] We did not use the full Muni Timetables for two reasons. Members tell us that it is very difficult to get a copy of the timetables, because they run out very quickly after they are issued; the frequencies advertised on all shelters and in the map available at drugstores are much more widely known. Also, to compare over 1000 data points with specific schedules in the timetables would have been extremely labor-intensive, and we simply did not have the resources.

[3] We also had some participants who monitored the system for a period of time, watching several vehicles go by and recording the intervals. Because these data points did not reflect a rider arriving at random at the stop, we set the "late?" variable to the *probability that a rider would be late*:

probability = (time waiting - advertised frequency) / time waiting

For the lines reported this way (J, 15, 30, 45, and several others) we averaged this probability into the total percentage of late arrivals.

[4]An extended comparison of the methodologies is found in the Technical Appendix.

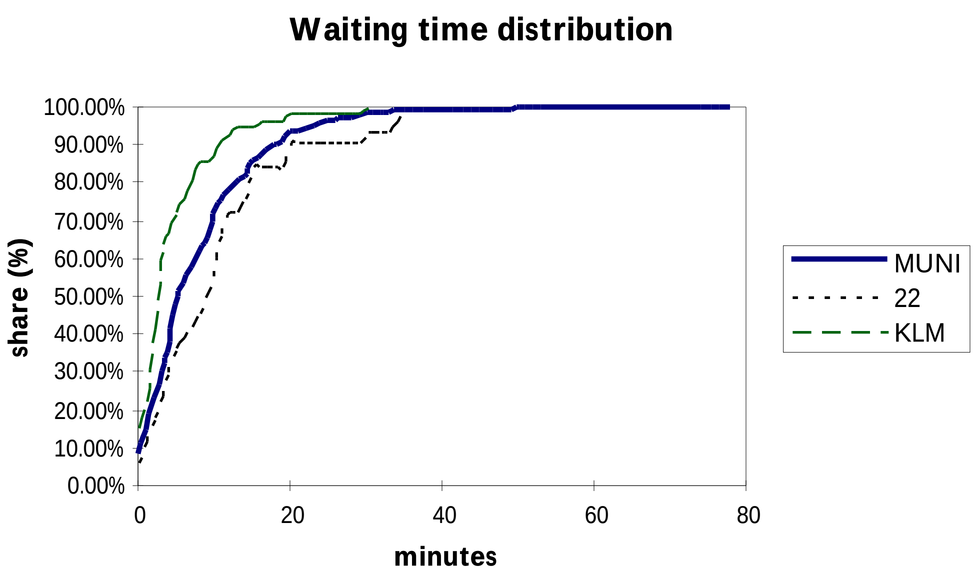
Key Findings

Overall system performance

Survey participants set out to test the hypothesis that Muni is unreliable. What we found was a Municipal Railway that consistently misses schedules and very frequently keeps riders waiting far longer than they should expect to wait. For the express commuter to the Financial District, this is bad enough; San Francisco is a place where one schedules 8 am meetings at one's peril because attendees may not make it on time. But for the rider who does not own a car and who must make several trips a day using multiple lines, perhaps to multiple jobs and errands across the city, this is a much more significant burden.

Overall, as stated above, the system ran late 25 percent of the time. Of the 1365 rides reported, 339 had waiting times longer than the frequency advertised in the system map. Depending on the line and time of day in question, this translated to waiting times of as little as ten minutes (for supposedly frequent lines like 1-California and 14-Mission) but as long as an hour for some less frequent lines. Long wait times were not isolated incidents; as the following chart shows, while the majority of riders in fact waited only a few minutes, 25 percent waited more than 11 minutes, and one rider in twenty waited more than 24 minutes. (The thick line in the chart is for all of Muni; above it is a line representing the relatively reliable KLM Muni Metro, and below it is a line representing the unreliable 22-Fillmore.)

Chart 1: Distribution of waiting time for MUNI and two selected lines



Put another way, 25 percent of riders waited longer than 100% of the advertised waiting time, but many of those riders waited much longer than that. We also analyzed the distribution of what we will call "normalized wait time", or waiting time as a percentage of advertised frequency. Seven percent of riders waited more than twice the advertised amount, and some waited as long as *six times* the advertised frequency. On average, our participants waited 82 percent of the advertised frequency. (A system running well would have this figure in the 50 to 60 percent range; see the Technical Appendix for a detailed discussion of this.)

Those who ride Muni every day are familiar with the crap-shoot nature of the system that these figures demonstrate. When they arrive at the bus or streetcar stop, while riders have a reasonable confidence that are *likely* to be on time, they have *no idea* if this particular time is their turn to be 45 minutes late, thanks to Muni. This makes it impossible to depend on the system for consistent service and pushes many riders into using automobiles or taxis (at much greater expense) just to avoid possible delays. To be able to rely on Muni, customers need to allow 75 to 90 minutes to cross a city that is only eight by eight miles.

Problem lines (and better lines)

On-time performance and related metrics differed markedly from line to line. While the system as a whole rated a C for its performance based on the standard teacher's grading scale, we found five lines that rated a D and three lines that rated an F. (This excludes lines for which we received fewer than twenty responses.) Lines graded F were 22-Fillmore, 19-Polk, and 1-California, and lines graded D were 15-Third, N-Judah, 44-O'Shaughnessy, F-Market, and 30-Stockton. The percentage late for each line is shown below. Also notable is the *average normalized waiting time* for these lines; riders on all three lines graded F wait *on average* substantially more than the advertised frequency, and no route graded D or F had riders waiting on average less than 80 percent of the scheduled time.

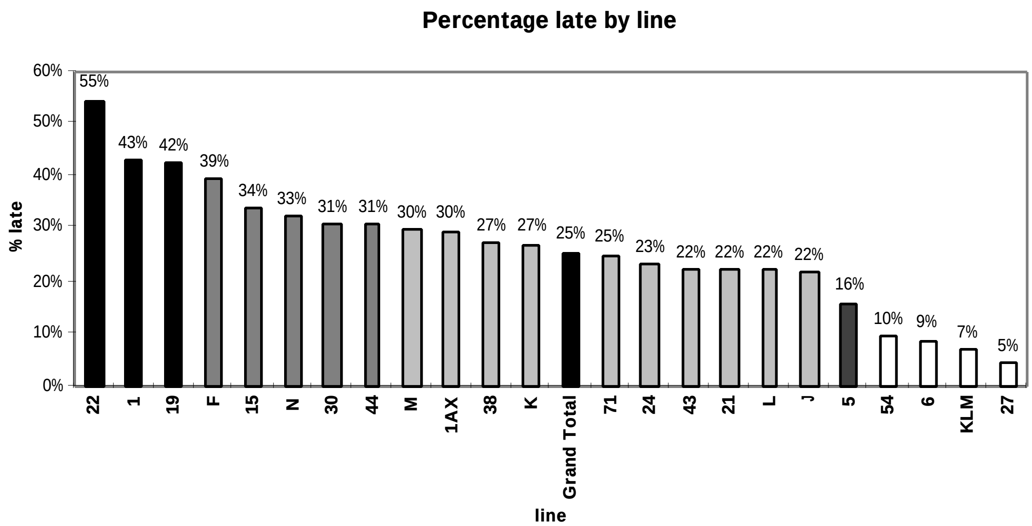
In sharp comparison, certain lines ran quite well from the rider's perspective. Four lines with sufficient data received a grade of A: 6-Parnassus (including some other Haight Street buses), 54-Felton, 27-Bryant, and the underground portion of the KLM Muni Metro. These routes were on time more than 90 percent of the time, and this was demonstrated also in normalized waiting time; the typical rider can expect to wait between 40 and 60 percent of the scheduled frequency on average, which proves to be about right for a line running well.[5] The one line graded B, 5-Fulton, also had the typical rider waiting a reasonable amount of time.

[5] Imagine a line that runs perfectly on a 10-minute schedule and a rider who arrives at perfectly random times to the stop. We would expect this rider to wait 1 minute 10 percent of the time, 2 minutes 10 percent of the time, and so on. This rider's average waiting time would be five minutes, fifty percent of the posted frequency.

Table 2: Ranking by line of on-time performance and normalized wait time (Excluding lines with fewer than 20 responses)

route	% late	grade	avg waiting time (min)	avg normalized wait time	StdDev of waiting time (min)	total responses
22	55%	F	10	116%	9	33
1	43%	F	8	130%	7	51
19	42%	F	11	119%	9	45
F	39%	D	9	102%	11	33
15	34%	D	8	93%	5	50
N	33%	D	6	84%	5	120
30	31%	D	4	97%	3	46
44	31%	D	12	92%	15	26
M	30%	C	8	76%	6	37
1AX	30%	C	10	98%	8	27
38	27%	C	4	65%	4	22
K	27%	C	8	79%	7	52
71	25%	C	10	77%	10	20
24	23%	C	7	70%	5	52
43	22%	C	9	75%	7	43
21	22%	C	6	71%	4	27
L	22%	C	7	85%	6	59
J	22%	C	10	88%	8	49
5	16%	B	5	75%	4	44
54	10%	A	12	60%	6	21
6	9%	A	7	54%	6	45
KLM	7%	A	4	41%	5	58
27	5%	A	7	47%	4	22
Grand Total	25%	C	8	82%	8	1365

Chart 2: Comparison of lines by on-time performance



Examples of lines

To illustrate how routes differ, we have identified examples of particularly poor performers (graded F), typical performers (graded D and C), and good performers (graded A).

22-Fillmore (Graded F): This line had the worst on-time performance among those with twenty or more responses. Not only was its on-time performance dismal (55 percent late); in addition, it had an *average normalized wait time* of 116 percent. This means that the typical rider, arriving at random throughout the advertised frequency (typically 8 to 15 minutes) waited 16 percent more than he should expect to wait *having just missed the previous bus.*

In addition, as the following chart shows, the 22-Fillmore rider has a 14 percent chance of waiting twice the frequency and a 5 percent chance of waiting *three times* the frequency. This translates into real wasted time; 25 percent of 22-Fillmore riders waited 14 minutes or more on this supposedly frequent route. Needless to say, this is a sign of a consistently unreliable bus line.

1-California (Graded F): This line also had very poor on-time performance despite (or perhaps because of) its extremely high advertised frequency. This was the line with the highest average normalized waiting time, 130%, with 43% of buses running late. This problem was particularly noticeable during rush hour: 55% of riders waited longer than they should, waiting 170% of the advertised frequency on average.

As with 22-Fillmore, a substantial fraction of 1-California riders waited much more than the advertised frequency. 14% of 1-California riders waited more than twice the advertised frequency, and one unfortunate rider waited almost seven times the advertised frequency, the longest relative wait recorded. Again, this was real wasted time: 10% of riders of this very frequent line waited 17 minutes or more, and one rider in twenty waited more than 23 minutes. (See Chart 1 above for an illustration of this in comparison to Muni as a whole.)

N-Judah (Graded D): This was the most widely reported line in the survey, with 120 responses. This line was a good example of the mediocre service common throughout Muni: it ran late 33 percent of the time, but riders did not on average wait more than the advertised frequency. (Average normalized waiting time was 84 percent, very close to the number for all of Muni.) Like 1-California, however, N-Judah was less reliable during rush hour, with 41 percent of rush hour streetcars late and average waits of 112 percent of advertised times.

Distribution of N-Judah waiting times was not as skewed as Lines 1 and 22, but it still left a bit to be desired. Only six percent of N riders waited longer than twice the advertised time, but these six percent found themselves waiting 19 minutes or more, and we did have one rider wait more than four times the posted interval. The latest twenty-five percent of N riders waited 120% of the advertised interval, which translated to ten minutes or more. This is in our opinion a significant if not egregious amount of wasted time.

24-Divisadero (Graded C): This was another example of a mediocre line; it performed slightly better than the Muni average, with 22% of riders reporting late runs, and with a mean normalized wait time of 70%. Like others, this line was later during rush hour than at other times (32% late), but it did not exhibit the long delays that lines discussed above had. Morning rush (35% late) was a particular problem here.

Distribution of 24 wait times was much better than on other lines. While 22 percent of riders waited longer than published frequency, no rider waited more than twice that interval. While 27 percent of riders waited more than 10 minutes, only 8 percent waited more than 15 minutes; because intervals are longer on 24 than on some other lines, this is not unreasonable. Perhaps this is due to better dispatching or scheduling than on some other lines; another thought is that the corridor traveled by the 24 has much less traffic than streets like Fillmore and Sacramento.

KLM-Muni Metro, between Embarcadero and West Portal (Graded A): Perhaps due to the controlled environment, the KLM Muni Metro (from Embarcadero to West Portal only) performed remarkably well. Only 7 percent of underground KLM trains were late, and this led to an average wait of 41% of the posted interval, the lowest among lines with sufficient data. Only 3% of KLM trains were late during rush hour, an excellent performance. Evenings were a problem, however; perhaps because the substitute service is not as reliable as the streetcars, evening KLMs were late 25% of the time.

With such a good on-time rating, one would expect a relatively good distribution of waiting times, and in fact that was the case. Only 10% of KLM riders waited more than 10 minutes, and while the longest normalized wait was 2.5 times the published interval, only three percent of KLM riders waited more than 1.5 times the advertised frequency. (Compare this with the 22, where 21% of riders waited more than 150% advertised frequency.) This does not include delays *after* the rider has boarded, however: many users commented that they experienced long delays waiting to arrive at Market Street stations in the morning rush hour. (See Chart 1 above for an illustration of this in comparison to Muni as a whole.)

6-Parnassus (Graded A): This line also did very well. Only 9 percent of 6-Parnassus vehicles (together with 7 and 66 buses reported as "Haight Street corridor") were late, and 6 riders waited 54 percent published frequency on average. The only time that 6 had some difficulty was on weekends, when 17 percent of buses were late, but even then the expected wait was only 62 percent of the published interval. (Rush hours were not significantly different from the total, 11% late.)

Distribution of 6 wait times was also not bad: nobody waited more than 2 times the advertised frequency, and 96% waited 1.2 x advertised or less. This translated to wait times of 16 minutes or less for 96 percent of riders. (One rider did wait 40 minutes, however.) Like the KLM, the 6 seems to be run reasonably well.

These data beg the question: why are there such differences among similar lines? Like 22-Fillmore and 1-California, 6-Parnassus runs down heavily traveled streets with slow traffic and much double-parking - yet it runs on time, even at rush hour, while 1 and 22 run very poorly. The KLM has a certain advantage since it runs in a tunnel, but even with the frequent breakdowns in the tunnels it seems to do better than most surface routes, particularly the other streetcars, which all scored C or D.

Performance by type of line and time of day

We also measured performance of different types of lines and different times of day in the aggregate in this survey. Rush hour vehicles were on average a bit later than vehicles at other times of the day (27% late rush hour vs. 23% late nonrush), and riders waited a bit more relative to the timetable (87% vs 77% normalized wait time), but these differences were not huge. Expresses mimicked the rush hour behavior, with 28% late and average waits at 92% of scheduled intervals, worse than the general population but not by a wide margin. Weekends and evenings in the aggregate were not much different from the total sample, but we did find that the buses run much better on Sundays (22% late) than on Saturdays (35% late).

There were also slight differences between electric and diesel coaches: older electrics ran 26 percent late and new electrics (on Routes 14 and 31) ran 30 percent late, while the diesel coaches ran 23% late on average. However, these differences were not nearly as significant as the differences between specific lines.

User comments

With the survey forms that were returned, we also received many comments on the quality of Muni service during the survey period and in general. Many riders reported trouble with the lines they were surveying:

- *Bus nearly ran into a car!*
- *I take Muni everywhere throughout the city for my work and I want to share with you that the Fillmore #22 line is usually and consistently not on any schedule. Need to wait forever. Sometimes I've waited 30 minutes or more for a #22, then 2 or 3 come... At times I've waited more than 40 minutes for the #37.*
- *Most Operators did NOT announce intersecting lines or major streets*
- *On 2/11 on my ride on the 22-Fillmore, the driver stopped the bus at Fillmore/O'Farrell so he could go buy some candy. I didn't get his id # because he wasn't wearing his coat.*
- *This survey does not capture the many times we give up. [Many comments said "gave up and walked."]*
- *[Line 1:] Lots of people waiting; 4 buses came at once!*

- *[Line 1:] two packed buses (7:42, 7:50); woman in next seat said she'd had to wait an hour!*
- *[Line 21:] terrible driver #3842; abrupt stops – passengers bracing to keep from falling forward*
- *[Line 22:] At least 90% of the time when I take the bus anywhere it is in conformity with the listed schedule. That however does not excuse the few times when the bus was grossly late. This lateness however can be attributed to many factors, the principal one I believe is the interference of traffic and stop lights on the bus route.*
- *[Line F:] full to capacity - 30 people left behind*
- *[Line M:] waited 25 min for M or 28, didn't come. M cancelled - no signs, no info. Approx 60 people as confused as me. Then a shuttle bus whistled past - no info!*

Some people noted an improvement in service, just during the survey period:

- *I am somewhat dismayed at how little I have waited during the survey period. Is there any reason to believe that Muni has beefed up service during this time (only to slip again after this week)?*
- *I am STUNNED at the improvement in service!!*
- *These were the best two weeks I have had on Muni in months. On Sunday 2/16 when the survey was over, it took 1.25 hours to get a 27 to Market in the morning. Then gave up waiting for a 21 and took an N car.*
- *[Line N:] It seems that during the two survey weeks, MUNI was on its best behavior... Waiting, however, did not seem to be the most aggravating part of my commute. Overcrowding, frequent back-ups in the Market Street tunnel, and dysfunctional elevators were the most frustrating... Since February 15 ... my wait times have become noticeably longer, and in many cases they've doubled. Overall MUNI is unreliable and inefficient ... No wonder there are too many cars in the city.*

These comments seem to confirm our assessment that Muni is consistently unreliable. Also, they serve as an important reminder of what can happen when Muni is held accountable; while we are skeptical that this survey was a cause of improvements in service, in general a closely watched system is a system that is "on its best behavior," as one rider commented. In the absence of viable competition, good supervision and third-party measuring may make a substantial difference.

Probable causes

This survey shows that Muni customers experience very poor service on some routes, but not on others. We believe that this clearly demonstrates that Muni has the ability to serve the city well; at the present time, it is not doing this effectively throughout the city. The following are some structural issues that we believe have an impact:

- *Budget:* Muni may simply not have the funding needed to provide as many runs as are shown in the schedule. In the past ten years, Muni's budget has been cut by a cumulative total of $30 million in real terms[6], and this may render it impossible to deliver service at the scheduled level of intervals on every line.

- *Traffic:* Several routes that we found performed particularly poorly (1, 22) are in high-traffic corridors not well-suited to mass transit. In addition, the Embarcadero and Central Freeways are down and this has significantly added to the congestion on the streets of San Francisco in the downtown corridor. Muni routes have not been re-timed to reflect the new traffic realities in many years. This creates tremendous pressure on the operators to meet their schedules given likely delays with loading passengers, frequent double parking, and so on.

- *Absenteeism:* The added stress on operators from not having sufficient recovery time from traffic delays, delays loading passengers, and having to carry a double load because of understaffing may result in further staff shortages due to absenteeism.

- *Supervision:* The 40% reduction in the number of street supervisors over the past decade has exacerbated the problem because without street supervisors, the system cannot be managed dynamically when problems occur. Also, of course, with effective street supervision operators can be held accountable for meeting schedules. A partial restoration of these positions has been promised but not yet seen.

- *Equipment failure:* Many riders reported long delays due to power failures, broken-down streetcars and buses, and similar issues. To Muni's credit, a capital program now in place to replace all Muni Metro streetcars seems to be making a positive difference.

Action required

Given the above, RESCUE Muni calls for the following from Muni and the city.

- *Scheduling:* Muni must stick to its schedules. If the schedules are unrealistic, and this survey suggests that they may be (particularly on very-frequent routes like 1-California), Muni should re-time major routes and publish a schedule it can meet. Together with this must be a much more serious effort by Muni to hold operators, supervisors, and managers responsible for the reliability of their lines; public disclosure of all internal performance statistics (discussed below) is a great way to achieve this.

- *Traffic:* This is clearly a major issue, particularly with widely-traveled routes such as 1 and 22. Muni needs to continue to work with SF Parking and Traffic to ensure priority enforcement against double-parking on congested downtown corridor streets, and it needs to push for more transit-only lanes on downtown streets. (The Market Street diamond lane seems to have made a positive difference.) Another useful step might be traffic-signal pre-emption, so that Muni vehicles can make the lights at intersections along congested roads.

- *Budget:* While RESCUE Muni is hesitant to call for a wholesale increase in taxpayer or farepayer funding for Muni at this stage, particular elements in the Muni budget may require some additional dollars. In particular, Muni should find a way to increase the number of street supervisors on patrol each day; a monitored system is in general a more reliable and accountable one.

---

[6] *1996 Muni Stakeholder Meetings: Summary of Proceedings,* January 30, 1997.

- *Communications:* This is one of Muni's biggest weaknesses today. When problems occur, riders are rarely told about them in a way that is useful for trip planning. Similarly, the public does not have access to Muni's internal reliability statistics. Muni needs to rectify this immediately; it must regularly publish data *in detail* on reliability, complaints, unexcused and excused operator and manager absences, and budget status. This would hold Muni accountable in a way that it is not today, and it would provide a means for the public to judge whether Muni's management is up to the task. In London, the Underground regularly publishes a "Customer Charter" that provides these details to the riding public; this would be an excellent model to follow.

Conclusion

The Muni Riders' Survey, conducted in February 1996 by RESCUE Muni, demonstrates the poor reliability of the San Francisco Municipal Railway and the wide variations in service quality from line to line. While Muni itself rated a C, with 25% of riders experiencing delays, several lines (rated F) had more than 40% of riders delayed during our survey. This is a major cause of customer frustration, as described in detail by many of our 97 participants.

We believe that this unreliability of Muni and of particular lines is a major reason for the decline in ridership in recent years, and we join the call for serious reforms in the way that Muni does business. In particular, we will insist on the real accountability that is expected for an organization supported by our fares and tax dollars but that is glaringly absent with Muni. This will help provide Muni the incentive it needs to become a world-class transit system again. San Franciscans may love to hate Muni, but we absolutely deserve better.

Technical Appendix

How do the results of the riders' survey compare to the published vehicle frequencies? Is Muni's performance consistent with its claims? This is not as easy a question to answer as it might sound, for two reasons.

First, the survey and published frequency are measuring two different things. Our survey records how long riders waited, whereas the published frequency, as given in the *Street & Transit Map*, specifies how many minutes elapse between buses. Only a rider who arrives just after the last bus departed should wait as long as the published frequency. Most riders should wait for a shorter time. Therefore, though it is simple and revealing to compare the riders' waiting times to the published frequency, this *understates* the percentage of buses that are late with respect to their schedule. If we are not careful, then Muni will seem to be running far better than other surveys have found.

Second, we also need to be careful about interpreting the published frequency. The *Street & Transit Map* does not specify the maximum time between vehicles, but rather the "average" time between vehicles. Even if Muni vehicles ran as scheduled, some fraction would naturally be late according to this measure.

Before we can properly assess the results of our survey, we need to take these two effects into account. We need to consider that riders arrive at some random time within the interval between vehicles. And we need to estimate how many vehicles would tend, even in the system ran flawlessly, to take longer than the average frequency to arrive.

The key question is, What fraction of riders would tend, even in a flawless system, to wait longer than the published "average" frequency? If this fraction is larger in practice than in theory, then Muni is failing to meet its claims.

The bottom line of our analysis, presented below, is that about 13 percent of riders in a flawless system would wait longer than the published frequency — compared to 25 percent of riders in our survey. Therefore, Muni is not flawless. The system is running late even when you take into account the inherent randomness of driving a bus or streetcar on city streets.

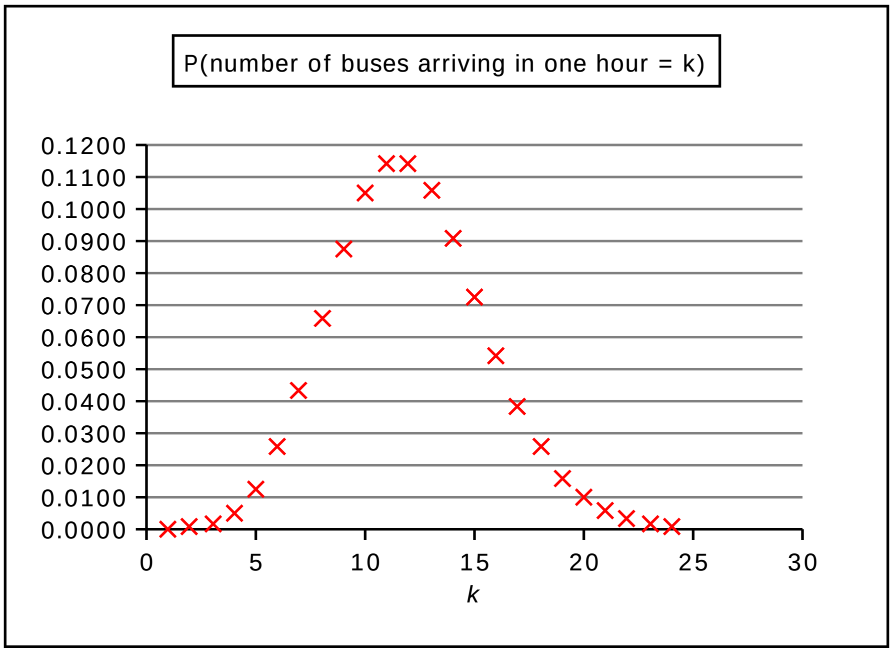
A model for vehicle arrival times

If traffic flow, red lights, and passenger loading and unloading were all perfectly predictable, then a vehicle would indeed arrive exactly once every interval. But this hardly ever happens in reality, except perhaps during Owl Service or in the Metro tunnel. Randomness in traffic and so on causes vehicles to speed up or slow down relative to their schedule.

To account for all the random influences on each vehicle would require an intricate computer model. In the end, such a model would not be terribly illuminating for our analysis of the survey, because conditions vary so widely from block to block and hour to hour. (On the other hand, computer models would help in the detailed planning of routes and evaluating the effect of reforms such as proof-of-payment fare collection. RESCUE Muni has started to have discussions with researchers who have developed such models for San Francisco.) In our case, there are enough different random influences on vehicles that we may turn to a simpler mathematical model of the aggregate effect. We have studied two possible mathematical models. Both reach almost the same answer to our above questions.

The standard model used by statisticians in situations of this sort assumes that the arrival of buses is a "Poisson process." The Poisson process involves vehicles that act independently of one another. Strictly speaking, this is not the case for Muni vehicles, because bunching causes a vehicle to depend on the performance of the one in front of it. But in this survey, it is seldom that riders take data on two successive vehicles; the vehicles they do measure are spaced apart and, therefore, approximately independent of one another. The alternative model we will consider later does not suffer from this limitation, but is more involved.

For example, suppose the schedule says a bus should arrive every 5 minutes on average. That is, we expect to see 12 buses in any given hour. It is also possible we will see fewer or more than 12 buses, according to the Poisson distribution:

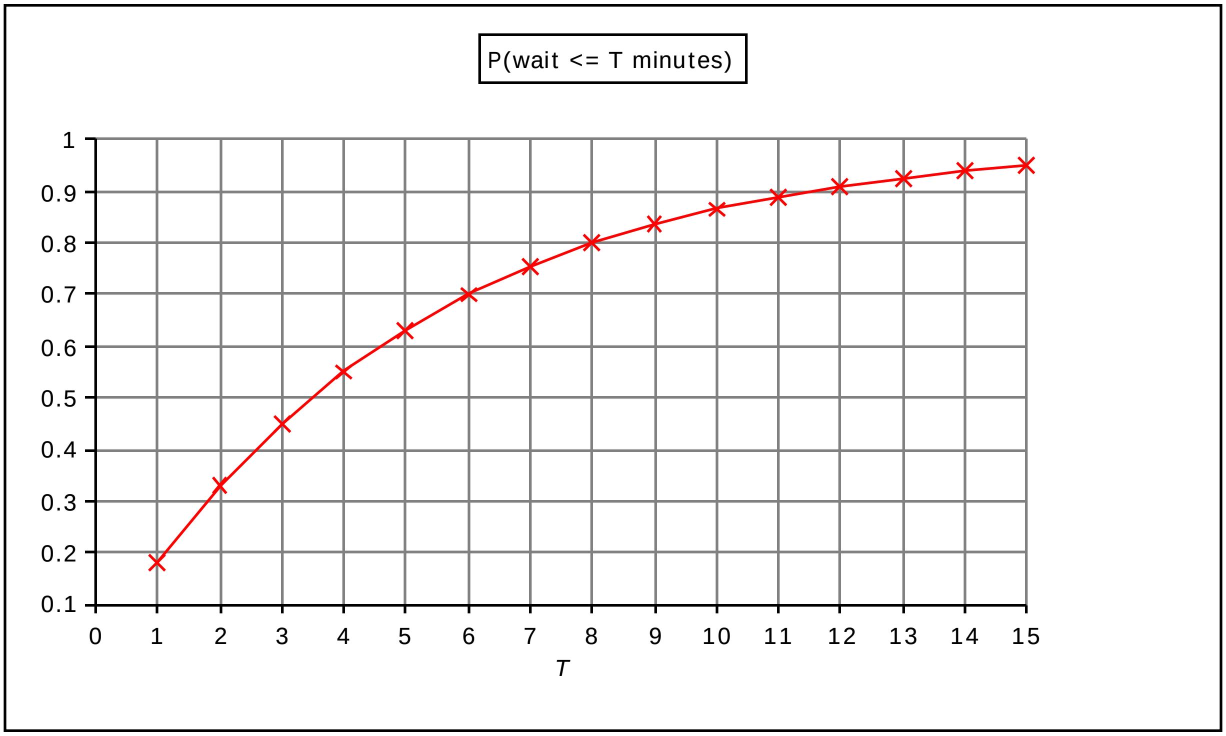


As this figure shows, the most likely scenario is 11 or 12 buses per hour, and the probability of other scenarios declines almost symmetrically. The probability that we get k buses per hour is

$$P(\text{number of buses in one hour} = k) = e^{-\lambda}\lambda^{k} \,/\, k! \qquad (1)$$

where λ is the expected number of arrivals per hour and $k!$ is the factorial of k.

In a Poisson process, the time between successive events — in our case, the arrival of a vehicle — is exponentially distributed. The exponential distribution differs from the more familiar bell curve, or normal distribution. For the exponential distribution with a mean of 5 minutes, the median interarrival time is only 3.5 minutes. Thus, we would expect that 50% of buses come less than 3.5 minutes after the previous bus, and 50% come after a longer wait — occasionally quite a bit longer. (We assume that Muni's published frequency refers to the mean.) The following graph summarizes the exponential distribution of interarrival times:



The chance that the interarrival time will be more than the published frequency is about 37%. The chance it will be twice the published frequency is about 14%; three times the frequency, about 5%. In general, for a published frequency of β minutes, the chance that the interarrival time is less than *t* minutes is

$$P(\text{interval} < t) = 1 - e^{-t/\beta} \qquad (2)$$

With the exponential distribution, most buses (63%) will arrive in under the published frequency, but the distribution has a long right tail, so it is possible that a very long time will elapse between buses. Conceivably, it is possible to wait half a day for a bus, although the probability is fortunately minuscule (you are more likely to be hit by a meteorite in the meantime). This is simply the nature of a random arrival process. The bus schedule does not claim a bus every 5 minutes — it only claims that one will arrive every 5 minutes on average.

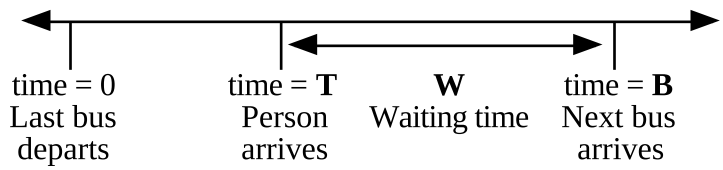
This does not mean that steps cannot be taken to reduce the impact of random processes. Subway tunnels, bus lanes, signal pre-emption, strict enforcement of parking ordinances, and proof-of-payment fare collection all help a transit system to run more predictably.

A model of rider waiting times

In order to compare the survey results to our theoretical model, we need to incorporate the rider. We assume that the rider arrives at the bus stop at a random time, independent of the arrival of the bus. In other words, we assume that riders do not synchronize their arrival at the stop to the bus schedule.

Some riders, especially those near the starting point of a line, are able to time their arrival at a stop. But as Muni patrons know, this is generally a forlorn hope. As the survey shows, the mean waiting time is comparable to the standard deviation of the waiting time. A bus arrives in 5 minutes ± 5 minutes — hardly a schedule that lends itself to predictability. This is, again, a natural consequence of the randomness of driving a vehicle on busy streets. In an exponential distribution, the mean and standard deviation are equal.

Without any knowledge of when the vehicle will come, the rider is as likely to arrive in the first minute after the last vehicle as in the second minute, as in the third minute, and so on, until the next vehicle comes. We can represent the situation as a time line:

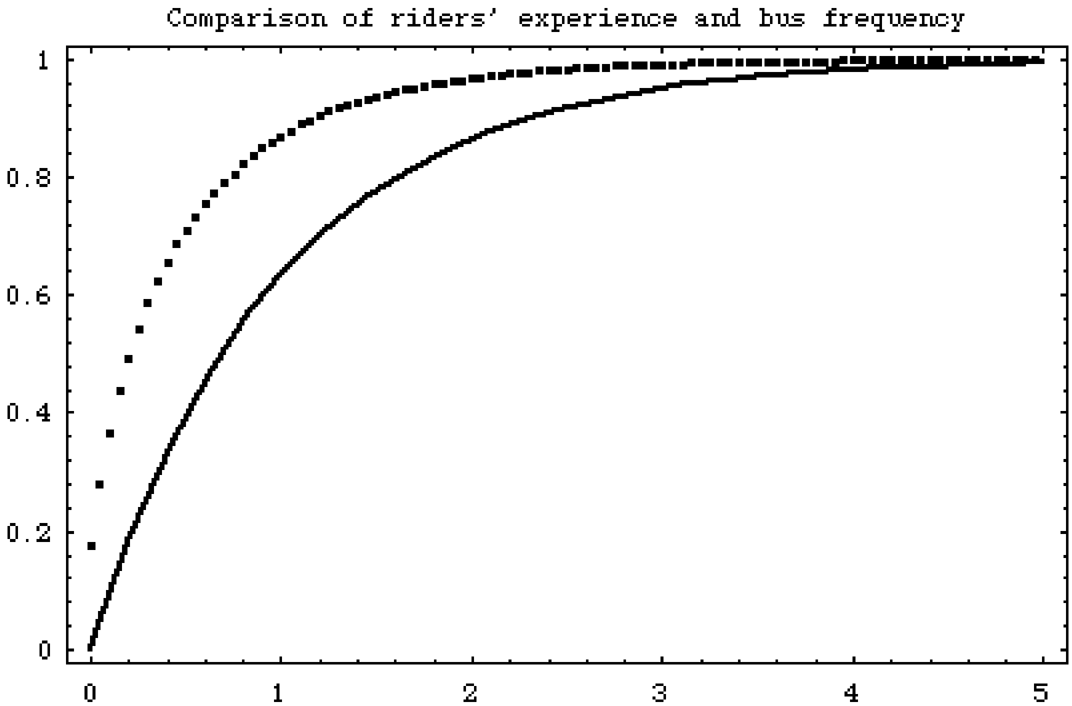


The rider's arrival time, T, is uniformly distributed on the time interval between vehicles, $(0,B]$. The rider measures a waiting time $W = B - T$.

We must combine this uniform distribution of rider arrival times with the exponential distribution of vehicle interarrival times. It turns out that there is no simple mathematical formula for the combination of these distributions. Instead, we have used a computer to perform a numerical simulation of the answer. The procedure is simple: We have generated a list of 10,000 exponentially distributed random numbers, $\{b_1, b_2, \ldots, b_{10000}\}$. These numbers represent the interarrival times of vehicles. For each of these numbers, b_i, we have generated a uniformly distributed random number, t_i, between 0 and b_i. The difference of b_i and t_i equals the waiting time, w_i. The result is, in essence, what we should have gotten if we had performed a 10,000-point survey on a hypothetical, ideal transit system.

Results of the Poisson-process model

We show the results in the following graph. The vertical axis is the probability; the horizontal axis is the multiple of the published vehicle frequency. The dotted line expresses the rider's point of view: the chance that the rider will wait less than a certain time. The solid line shows the system's point of view: the chance that the vehicle interarrival time will be less than a certain time.



Here are the same results in table form:

multiple of published "average" frequency	chance of a shorter rider waiting time	chance of a shorter vehicle interarrival time
0.5	71%	39%
1.0	87%	63%
1.5	94%	78%
2.0	97%	86%
2.5	98%	92%
3.0	99%	95%

For example, suppose the published frequency is one bus every 10 minutes. There is a 71% percent chance the bus will come by the time the rider has waited 5 minutes, and an 87% chance that the bus will come by the time the rider has waited 10 minutes. Only 1 in 100 riders should wait 30 minutes or longer.

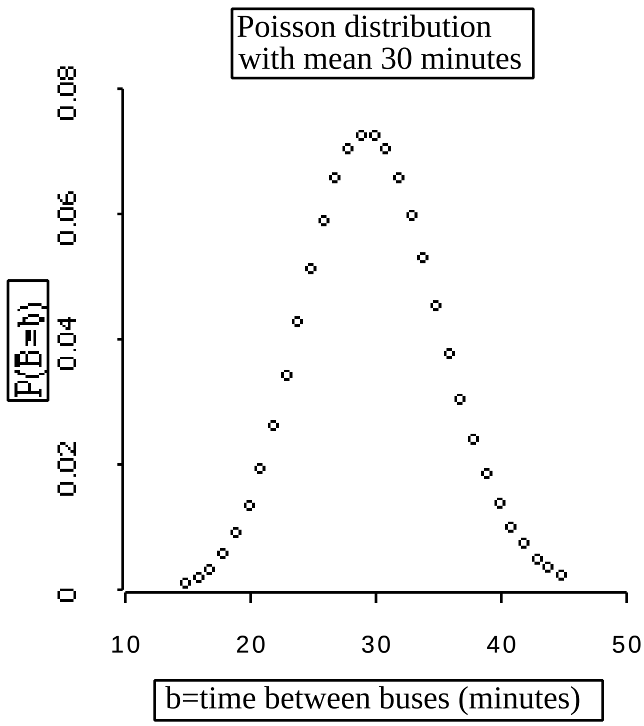
From this we see that, in a flawless system, only 13% of riders would wait longer than a full schedule interval. All Muni lines that meet this criterion in our survey earn an 'A' grade. Most do not. In the system as a whole, 25% of riders wait longer than a full schedule interval — nearly twice as many as should. The system is failing to attain its advertised performance.

Previous surveys by Muni and by the *Examiner* and RESCUE Muni (see the Dec. 2, 1996 *Examiner*) have measured the interarrival time, rather than the time that a rider can expect to wait. These surveys have found that the interarrival time exceeds the published frequency in roughly 50% of cases. According to the above table, this percentage would be 37% if the system were flawless. Clearly it is not flawless. The conclusion of these previous surveys is the same as for the present survey.

An alternative model of vehicle arrival times

As we mentioned above, the Poisson process is not a rigorously accurate model for Muni service, although it is probably sufficient for our purposes. What alternatives might we consider?

Rather than assume that the *number* of vehicles in a given time period is Poisson, let us assume that the *time* between vehicles is Poisson. This is a subtle but potentially significant change from the above Poisson-process model. Retaining the Poisson distribution in some capacity is reasonable because this distribution occurs naturally in a wide variety of phenomena. The Poisson distribution is almost symmetric about its mean, somewhat like the normal distribution, although it has a long right tail (going to infinity) and short left tail (ending at zero) — both of which are desirable for estimating vehicle interarrival time.



The Poisson distribution has two other useful properties for our purposes. First, it is discrete; that is, it deals in whole numbers, rather than continuous, decimal numbers. This matches our data, which measure time discretely (in minutes). Second, the Poisson distribution requires only a single parameter: the mean. This is the only parameter available to us from Muni's published frequency table. Other possible distributions, such as the negative binomial and the gamma distributions, would require two parameters. This, in turn, would require us to make additional assumptions about the functioning of the system.

Under the alternative model, the conditional probability of the waiting time of a person, given the arrival time of the vehicle, is

$$P(W = w \mid B = b) = 1/b \tag{3}$$

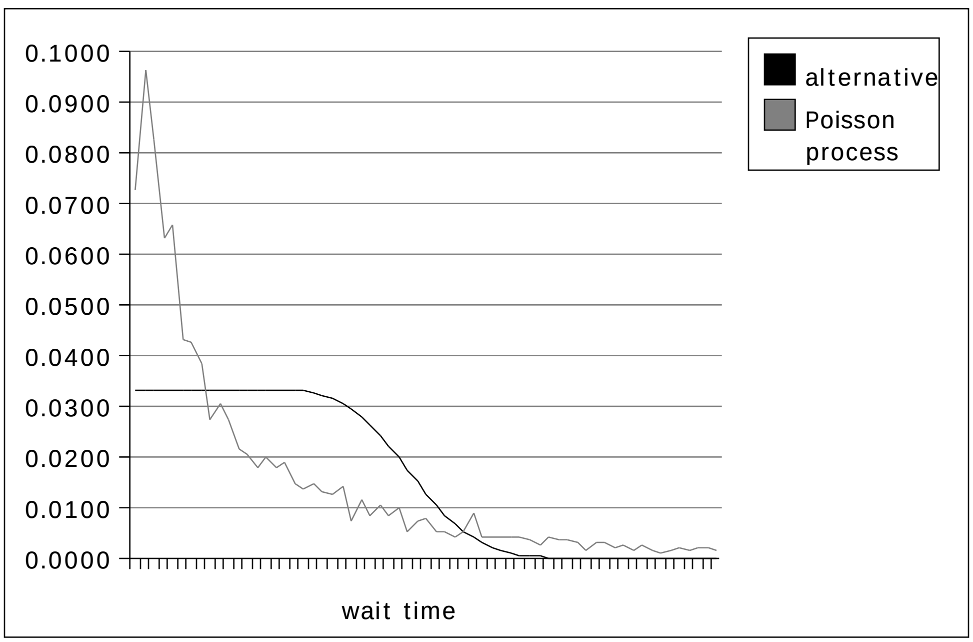
According to the Poisson distribution,

$$P(B = b - 1) = e^{-\lambda}\lambda^{b} \; / \; b! \tag{4}$$

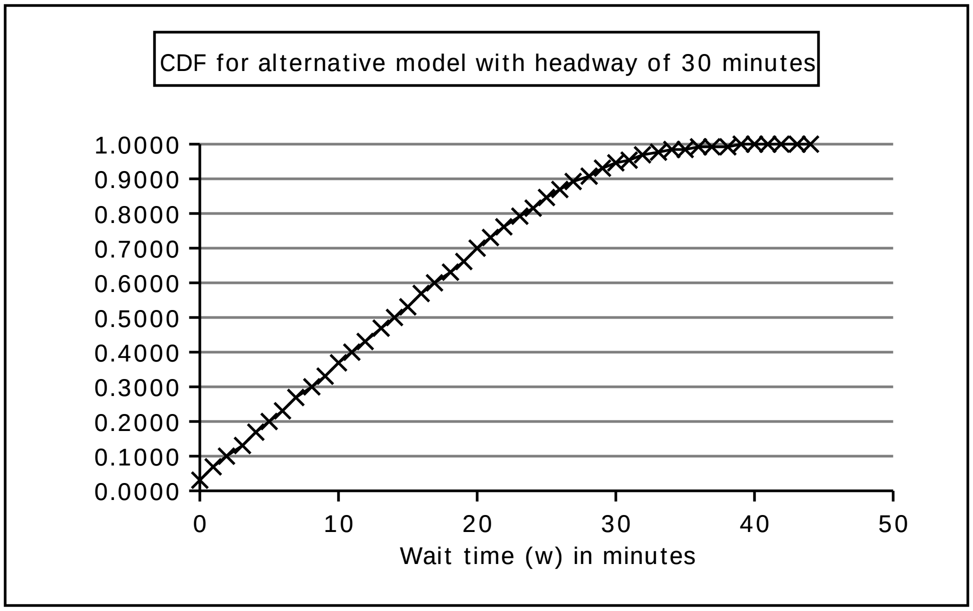
Equation (4) is expressed as $P(B = b - 1)$ in order to ensure that the vehicle does not arrive in the same minute as the previous one. This is purely for mathematical convenience; several vehicles can arrive at once, but the rider can only board one of them. By the Law of Total Probability, we derive a formula for the probability that the rider must wait *w* minutes:

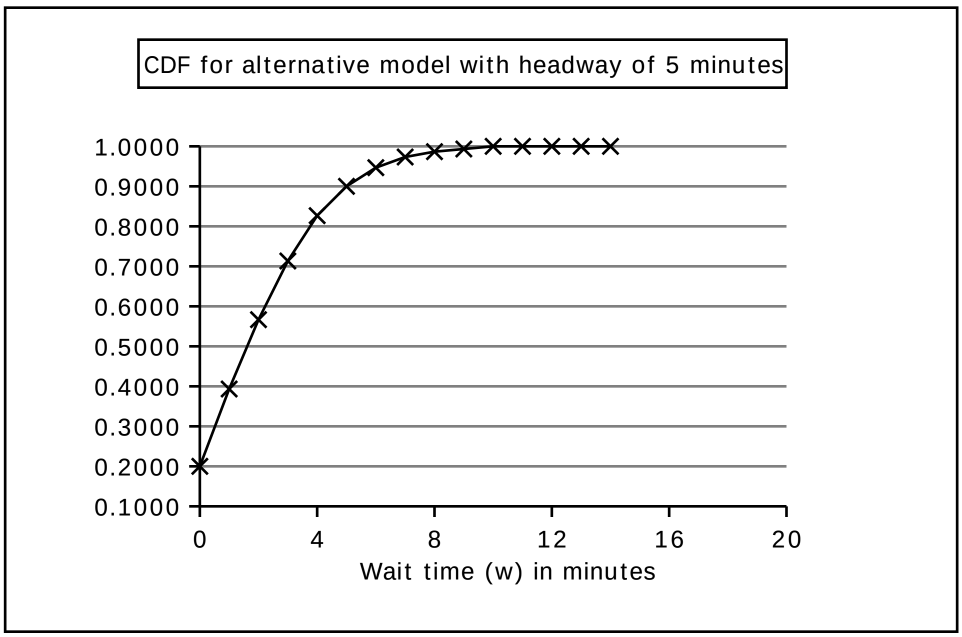
$$P(W = w) = \frac{1}{\lambda}\left(1 - \sum_{b=0}^{w} \lambda^{b} e^{-\lambda} / b!\right) \tag{5}$$

This formula is plotted as the solid line in the following graph. The vertical axis is probability; the horizontal axis is waiting time. The dotted line is the equivalent function for the Poisson process, based on the simulation described above. In the Poisson-process model, the probability declines steadily for longer waiting times. In the alternative model, the probability remains uniform from 0 minutes to half the scheduled interarrival time (15 minutes in this graph). After that, the probability of waiting longer falls off sharply.



Unlike the Poisson-process model, the alternative model has slightly different curves depending on the exact value of the mean interarrival time. The following two graphs show the probability (vertical axis) of waiting less than a given number of minutes (horizontal axis) for two sample schedule frequencies:





Here is the same information in table form. The table lists the chance that a rider will have a shorter waiting time than some multiple of the published frequency:

multiple of published "average" frequency	alternative model (5-minute frequency)	alternative model (10-minute frequency)	alternative model (20-minute frequency)	alternative model (30-minute frequency)	Poisson-process model
0.5	71%	59%	55%	53%	71%
1.0	90%	92%	93%	96%	87%
1.5	95%	97%	100%	100%	94%
2.0	100%	100%	100%	100%	97%
2.5	100%	100%	100%	100%	98%
3.0	100%	100%	100%	100%	99%

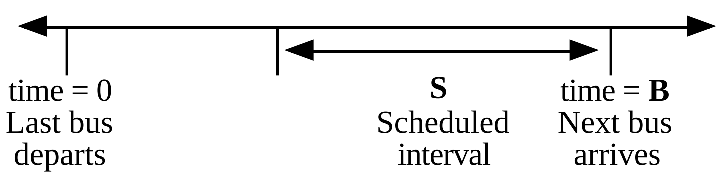
Compared to the alternative model, Muni performs slightly worse than it does compared to the Poisson-process model. At most 10% of riders wait longer than one schedule interval — compared to the 25% that the survey found. Only 1 in 1000 riders should wait longer than two schedule intervals. In practice, such long waits are far more common. This implies that the long waits are a consequence of breakdowns in the system (mechanical problems, absenteeism, poor dispatching) rather than a natural effect of randomness.

As we have seen, modeling the time between buses is, at best, inexact. The only convincing test of our simple models would be a comparison to a detailed computer simulation, a process we are beginning to undertake. The models do support the anecdotal experience of riders: The occasional slight delay is understandable, but Muni's delays are longer and more frequent than they should be.

Additional issues, 1: 'Monitoring' data

Of the 1,375 data points in the survey, 46 were "monitoring" data. For these data points, the rider, instead of arriving randomly at a bus stop and getting on the first bus, stood at the stop and recorded the times that buses passed by. These data are not quite comparable to the majority of survey data, because they consistently record the longest possible wait for the rider (B, rather than W). If we simply averaged these monitoring data into the survey results, those lines would appear to perform worse than they actually did.

Instead, we perform a correction that allows these data to be directly compared to the normal survey data. Suppose that, instead of monitoring the route, the rider arrived at some random time uniformly distributed on the interval between vehicles. This interval is $(0,B]$ on the following time line:



Here, S is the published frequency. The time line consists of two segments. If the rider arrived within S minutes of the bus arrival time, he or she would have recorded the vehicle as "on-time." Otherwise he or she would have recorded it as "late." The probability of the latter is

$$P(w > s) = (b - s) \ / \ b \tag{6}$$

where b is the time that the monitor recorded (that is, the actual interarrival time).

This probability given by Equation (6) fits naturally into our scheme for interpreting the other data points. For the majority of points, we assign a 0 if rider waited less than a schedule interval and 1 otherwise (that is, a Boolean variable). For the monitoring data, this number is between 0 and 1. Any given bus in these data has a certain chance of having been recorded as late.

The same formula may also be used to transform the results of earlier surveys — such as the *Examiner*/RESCUE Muni survey described by Erin McCormick in the Dec. 2, 1996 *Examiner* — into the same form used by our latest survey.

Additional issues, 2: 'Normalizing' the data

The collected data is complex: multiple routes, different bus stops on the same route, different times of the day, different days of the week. One way to analyze an entire line — or the entire system — with different interarrival times is to normalize the waiting times by the expected frequency of service. In other words, we divide the waiting times by the published frequency. For an exponential distribution, the process is statistically kosher. It does not change the underlying distribution of vehicle arrival times, but does allow us to make statements about the average performance of a line or the system.