Financial Services WORKING PAPER

0397

Reliability Analysis of the Federal Reserve Automated Payments System

by Apostolos Burnetas,
Gregory Reynolds,
and James B. Thomson

FEDERAL RESERVE

FINANCIAL SERVICES

Financial Services Working Paper Series No. 03-97

RELIABILITY ANALYSIS OF THE FEDERAL RESERVE AUTOMATED PAYMENTS SYSTEMS

by Apostolos Burnetas, Gregory Reynolds, and James B. Thomson

Apostolos Burnetas is an assistant professor of operations research at the Weatherhead School of Management, Case Western Reserve University. Gregory Reynolds is at Hewitt Associates, LCC, Chicago, Illinois. James B. Thomson is director of financial services research at the Federal Reserve Bank of Cleveland. The authors are grateful to the staff at Federal Reserve Automation Services (FRAS) for providing information and helpful comments on the design and operation of the Federal Reserve's automation infrastructure.

November 1997

Abstract

This paper proposes an analytic framework for the reliability assessment of the automated payments systems used by the Federal Reserve Banks. The failure/recovery behavior of the system currently in operation is modeled as a continuous-time Markov process with varying levels of detail, and the availability is calculated for a wide range of component failure frequencies. Furthermore, alternative system configurations are proposed and analyzed.

Reliability Analysis of the Federal Reserve Automated Payments Systems[1]

Apostolos Burnetas[2], Gregory Reynolds[3] and James B. Thomson[4]

November 12, 1997 - Comments Welcome

Abstract

This paper proposes an analytic framework for the reliability assessment of the automated payments systems used by the Federal Reserve Banks. The failure/recovery behavior of the system currently in operation is modeled as a continuous time Markov process with varying levels of detail, and the availability is calculated for a wide range of component failure frequencies. Furthermore, alternative system configurations are proposed and analyzed.

---

[1] We are grateful to the staff at Federal Reserve Automation Services (FRAS) for providing us with information and helpful comments on the design and operation of the Federal Reserve's automation infrastructure. The views presented in this paper are the authors' and do not necessarily reflect those of the Federal Reserve Bank of Cleveland, the Board of Governors of the Federal Reserve System, or the members of the Federal Reserve's Financial Services Policy Committee.

[2] Department of Operations Research and Operations Management, Weatherhead School of Management, Case Western Reserve University, Cleveland OH, (216) 368-4778, atb4@po.cwru.edu

[3] Hewitt Associates, LCC, Chicago, IL.

[4] Financial Services Research Group, Federal Reserve Bank of Cleveland, Cleveland, OH, (216) 579-3022, jb.thomson@clev.frb.org

Introduction

Federal Reserve Automation Services (FRAS) embodies the consolidation of the mainframe computing and communications operations of the twelve Federal Reserve Banks (FRBs). FRAS was set up to capitalize on the efficiencies that could be gained by consolidating the back offices of the regional banks. Equally important was the opportunity FRAS created to build more reliability into the automation infrastructure over which the FRBs operate their electronic transfer systems (Fedwire and Fed ACH).

Fedwire is one of many payments operations run by the FRBs. It is an electronic funds transfer system that links the twelve Federal Reserve Banks with more than 11,000 depository institutions nationwide. Fedwire is used by these institutions to clear and settle more than a quarter million transactions per business day, with an aggregate value of approximately $800 billion. These numbers are enormous when compared with the dollar volume of any other payment operation run by the Federal Reserve. Because of the size of the operations involved and their role in the economy, the reliability of the system handling these transactions is critical.

The system currently under operation is structured with a high degree of redundancy and very detailed backup policies and contingency plans, to ensure extremely reliable operation. However, the marginal costs associated with a high degree of reliability are typically very high. The purpose of this paper is to develop a general mathematical framework to assess the reliability of the existing system as well as less costly alternatives in configurations and backup policies.

The modeling and analysis methodology is based on representing the system operation as a continuous time Markov process in steady state. The main assumption of such a model is that the failure times for the individual system components are independent random variables following exponential

distributions. This is an accurate approximation for highly reliable computer systems, which are typically upgraded by newer technologies before the failure rate starts increasing.[5]

The costs associated with the reliability of FRAS are not considered in this paper, for the main reason that the operations involved are highly reliable and cost data associated with failures are scarce. The approach taken here is to explore the tradeoffs in reliability between the existing and alternative system configurations.[6]

The paper is organized as follows. The FRAS infrastructure is outlined in Section 2. Section 3 presents a statistical reliability analysis of the existing configuration. The availability of Fedwire and Fed-ACH is computed as a function of expected site failure rates. The sensitivity of these results to assumptions about individual machine failure rates and the reliability of the communications system is explored. Section 4 provides an analysis of alternative configurations for FRAS using an analogous modeling approach. Conclusions and recommendations are presented in section 5.

2. System Description

2.1 Applications

The FRAS system uses multiple mainframe computers located at three different sites to operate Fedwire.[7] Along with Fedwire, these mainframes are also used for several other applications; any study or recommendation for the system design and operation should consider the impact on them as well. The two systems that were considered the most critical after Fedwire were Automated Clearing House (ACH) and the software used to manage the communications configuration, called the CMC. ACH is a fully electronic batch-processing system through which value-dated payments are processed. Although ACH's dollar

---

[5] For a discussion on the appropriateness of the exponential distribution for modeling failures of reliability systems, see Keilson (1987). General methods for reliability analysis of systems consisting of renewable components are contained in Gertsbakh (1989).

[6] For an analysis of the economies of scale and the effect of technology changes on the FRAS operating costs, see Bauer and Ferrier (1996).

[7] A more detailed description of the Federal Reserve's role in the payments system, and in particular its automated payments systems, is found in United States General Accounting Office (1997).

volume is dwarfed by Fedwire, ACH handles many more transactions and therefore uses a larger amount of computing resources than Fedwire; ACH transactions are of a smaller average dollar value than those handled by Fedwire. The CMC manages communications sessions for internal (FRB) and external users of the FRAS systems; it is a critical component of the overall system reliability.[8]

2.2 Configuration/Backup Policies

As noted, FRAS operates from three geographically distinct sites. These sites, henceforth referred to as CC1, CC2, and CC3, were selected to minimize the impact of natural and other disasters on the operation of the system. CC1 is the main processing site for both Fedwire and ACH, while CC2 and CC3 serve as standby and recovery sites. The backup/recovery policies implemented by FRAS are detailed hierarchical contingency plans that specify : (a) the primary processing units for each application, (b) the "hot" and "cold" standby facilities, and (c) the order of transfer of the operations in case one or more processing units become inoperable. These events are referred to as failures in the sequel.

The following is a brief description of the main recovery policies for Fedwire and ACH.

2.2.1 Fedwire

CC1 is the main processing site for Fedwire, while sites CC2 and CC3 serve as hot and cold standby facilities, respectively. Transactions from depository institutions are received by CC1; they are then processed and recorded in a mainframe in that site. An identical image of each transaction is also executed and archived on backup systems operating both at CC1 and at CC2 (hot standby). In this mode of normal operation, the Fedwire application in site CC3 is not active (cold standby).

If the Fedwire application on the primary system at CC1 becomes inoperable, then local recovery occurs and the backup at CC1 becomes the primary processing unit. The recovery operation within site CC1 is practically instantaneous. This operation will be referred to as local recovery.

If both systems at CC1 fail, then the hot standby unit at CC2 becomes the primary processing unit. The targeted recovery time (amount of time until transactions can be processed again) in this case is less

[8] A detailed description of the CMC system is not provided herein because of security concerns.

than 30 minutes. In the event of failure of CC1 and transfer of main processing to CC2, the cold standby unit at site CC3 is switched to hot. This operation requires approximately six hours and it involves preparing site CC3 to replicate the transactions in real time.

Similarly, if both systems at CC1 and the primary backup system at CC2 fail, then transactions are handled by a similarly configured backup system at CC2. If this backup system also fails, while all the previous are still inoperable, then transactions are handled at CC3.

Regarding the repair and reinstitution of failed units the typical repair time of a failed mainframe unit is on the order of one day, depending on the type of damage and the availability of spare parts. Under normal circumstances, when a mainframe fails and is subsequently repaired, it is not brought back online until either the weekend or in extreme cases during an evening of a weekday. This is done in order to avoid potentially troublesome setup operations during normal operating hours. However, the setup of a repaired unit can be expedited if other units have failed in the meantime.

The system described above meets very high reliability standards. Not only would five systems need to fail in order to render the system inoperable, but also they must all fail in a short time so that no unit can be repaired before the last one fails. This makes the failure of the entire Fedwire operation an extremely rare event.

2.2.2 ACH

ACH is handled in a similar way to Fedwire. Site CC1 is the main processing center, a mainframe computer as the main processor and a second system as the onsite hot standby unit. However, unlike Fedwire, ACH operates with only one off-site contingency processor at CC3 as hot standby. The contingency plans for recovery are analogous to Fedwire.

3. Analysis of the Current System.

In this section we analyze the reliability of the existing system configuration. The reliability criterion used is the availability or average up time of Fedwire and ACH. That is, reliability is defined as the long-run expected fraction of time that each application is operative. Other related reliability criteria,

such as expected time to failure and probability of uninterrupted operation over a specified period can be used; however, the availability is considered the most important measure.[9]

The principal difficulty with the statistical analysis is the insufficiency of historical data on individual computer as well as entire site failures. The lack of data can be attributed mainly to the high reliability of the equipment involved in the operation, which makes failures very rare. It is indicative that an entire site failure has never occurred during FRAS's five-year existence. We confront this problem by analyzing the system for a reasonably wide range of values of the missing failure parameters. This approach makes it possible to estimate the relative importance of the primary and standby computers in the overall system reliability, as well as the sensitivity of the estimates to changes in the parameters.

3.1 Modeling Approach.

To assess availability, the system is modeled as a continuous time Markov process, a powerful mathematical model for analyzing the behavior of probabilistic dynamic systems (c.f. Kao (1997), Karlin and Taylor (1981)). The main elements of the model are described below.

We assume that the operational state of the system at any point in time can be fully described by a number of mathematical quantities, collectively referred to as the *state*. The necessary state information for each application (Fedwire and ACH) depends on the operational characteristics of the underlying process. In general the state information at any time must be sufficiently detailed to determine exactly all events that may possibly alter the state of the system in the immediate future (*state transitions*), as well as the probability distribution of the times until these events occur. However, there is a compromise between very detailed state modeling and the computational complexity of the resulting model.

For the system under analysis, there are several alternatives regarding the level of detail in modeling. The simplest choice is to model the entire system as a single unit that can be operational or not.

---

[9] An important feature of the overall reliability design is that, according to the recovery plans in use, if two mainframes fail at the same site, then all applications at that site are transferred to the other sites depending on their individual recovery rules and the availability of standby units at those sites.

In this case the system can be in one of only two possible states. However, the resulting model is very poor in representing the actual system operation. For example, if at some point in time the system is not operational, this might have been caused by a catastrophic failure of all the mainframes or by a delay in data transfer during a mainframe switch over. Although the consequences of these two situations on the future state of the system are very different, they cannot be captured by the 0-1-type state information.

At the other extreme, it is possible to include the detailed operating condition of each component of the mainframe computers and communication lines in the system state. In this case the model is very realistic, because the available information allows modeling of the entire variety of failure, repair, recovery and setup contingencies. The difficulty with this approach is that the resulting number of states is typically too large for analytical treatment.

This paper follows an intermediate path. The state is modeled as the operating condition of each mainframe within each processing site. The state of each mainframe can be any one of the following: normal operation, under repair, or under setup. When a mainframe is in the normal operation state, it may be either processing transactions or be in cold standby, according to the recovery policy specifications. When it is under setup, it has been successfully repaired from a previous failure and the process of connecting it to the system is under way. The time under setup also includes the period that a newly repaired mainframe may be idle, until a convenient time for connection arrives (weekend or end of day).

Initially, the communications connections are not included in the system state. Instead, the lines are assumed always operational. This assumption allows isolating the effect of the mainframe layout and operation on the system reliability. Furthermore, it is not very restrictive, because the communications lines are typically very reliable and there are always more than one possible routing paths for transmitting a transaction from a bank sending or receiving payments to the processing site. The joint effect of the mainframe computers and communications connections on the reliability is examined in a subsequent model, where communications are assumed prone to failure. Because the size of this extended model is

prohibitive for analytic solution, a simplified version is formulated, that provides a lower bound (worst case analysis) for availability.

Regarding the equipment behavior, two types of failures have been modeled. The first is site failures, i.e., failures that render an entire site inoperable and may be caused by catastrophic events such as fires, floods, etc.. The second type is individual mainframe failures that only render one mainframe inoperable and are typically caused by equipment malfunction. In this Markov process framework it is a standard assumption that failures of different types and different components occur independently of each other and the random time until failure of any unit follows exponential distribution. A significant implication of the exponential distribution assumption is that the probability of failure of an operating unit (mainframe or entire site) within a short time interval is independent of the total time that the unit has been previously in operation (constant failure rate, or "as good as new" assumption). This assumption is reasonable for modeling electronic or other equipment that are very reliable during their useful lifetime and their failures are caused by completely random (rare) events, as is the case in the present system.

It remains to discuss the modeling of repair and setup times. These are also assumed random and following exponential distribution with appropriate rates of repair and setup. The exponential distribution assumption is less plausible in this case than in the case of failures, because the repair and setup operations are more predictable than failure times. However, they are several orders of magnitude smaller as well, so that the assumption does not introduce significant errors in the results. Based on actual experience as well as maintenance contracts with the manufacturer, the repair times for a mainframe typically vary around a period of one to two days. Repair times for an entire site have not been experienced because entire site failures have never occurred, although duration of the same order of magnitude is reasonable to assume.

The randomness in the setup time modeling accounts for the fact that a repaired mainframe becomes operational at a random instant, but it is typically not reconnected into the system until the end of the week, to avoid unforeseen problems of reconnecting during system operation. The expected time required for the setup, as it is used in the model, depends on whether the repaired unit is needed

immediately or not because other mainframes are also under repair. If this is the case, an expedited setup process is assumed, with expected setup time in the order of several hours. Otherwise, the expected setup time is several days.

Finally, as is mentioned above, there is a short delay in changing from one mainframe to another depending on which mainframe fails and which mainframe will be taking over. For Fedwire this time is usually in the order of 15 minutes with one exception. Namely, when site CC1 fails and site CC2 becomes the primary processing site, then CC3 is switched from cold to hot standby operation within six hours. If site CC2 also fails during this six-hour period, then it will not be possible to switch primary processing to CC3 within 15 minutes. However, if site failures occur independently, then the probability of two sites failing within six hours is one order of magnitude smaller than that of a single failure; therefore, it can be safely assumed that the changeover time is always on the order of 15 minutes.

3.2 Analysis of Results: Existing System - Reliable Communications.

As discussed in the previous section, the quantitative analysis is presented for a range of values of the failure rates, in order to alleviate the uncertainty regarding these parameters.

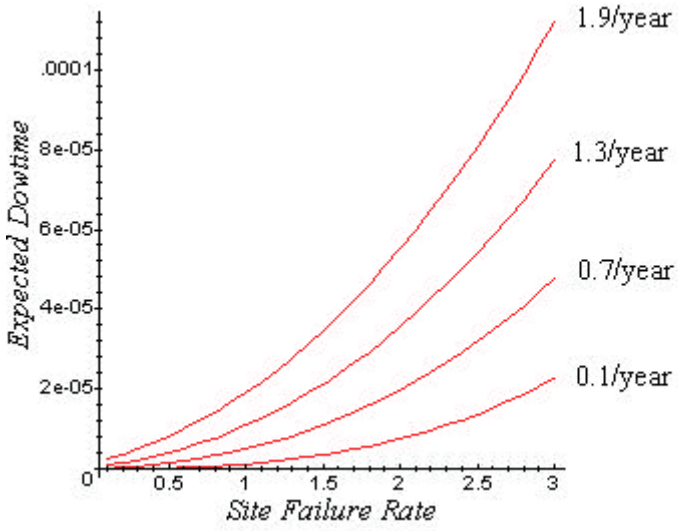

Figure 1: Expected Downtime of Fedwire

Figure 1 presents the expected downtime, d_F, of Fedwire, i.e., the proportion of time in the long run that Fedwire is inoperable, as a function of the mainframe and site failure rates. The availability, or expected "up-time", u_F, is given by $u_F = 1 - d_F$. In this model the communication lines are reliable, the failure rates of all mainframe computers are equal to λ_M, and the failure rates of all sites equal to λ_S. Parameters λ_S and λ_M denote the expected number of site failures and machine failures per year, respectively. Parameter λ_S varies from 0.1 to 3 and λ_M from 0.1 to 2.0 failures per year. Each curve represents the expected downtime of Fedwire, d_F, as a function of λ_S, for a fixed value of λ_M. Different curves correspond to different λ_M values. The numerical computations were performed for all pairs of λ_S and λ_M values in steps of 0.1. Although only the curves corresponding to λ_M =0.1, 0.7, 1.3, and 1.9 expected failures per year are included in Figure 1, the numerical results for other values are similar.

The data in Figure 1 indicate that the operation of Fedwire is extremely reliable. The expected downtime is always below 0.01% (1 hour per 10,000 hours of operation), even for unrealistically high values of λ_M and λ_S. This is mainly due to the high degree of backup redundancy in the system configuration.

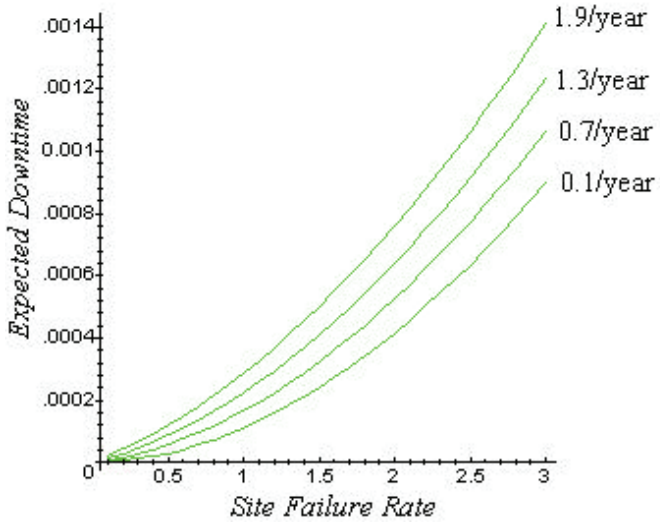

Figure 2: Expected Downtime of ACH

Figure 2 presents the corresponding results for the expected downtime of ACH, d_A, as a function of the machine and site failure rate parameters, varying in the same ranges as for Fedwire. In relation to Fedwire expected downtime, d_A is approximately 10 times higher than d_F. However, in absolute terms, the reliability of ACH is also very high, the downtime never exceeding 0.1% and generally varying around 0.03% for more realistic λ_S values of about 1 failure per year.

The comparison between the expected downtime in Figures 1 and 2 is also interesting to the extent that ACH can be considered an alternative system design for Fedwire. Therefore, it provides an indication of the tradeoff in reliability between the three-processing-site, five-mainframe configuration of Fedwire and the less costly two-processing-site, three-mainframe configuration of ACH. It follows from the above discussion that the difference in reliability between the two systems is mainly attributed to the existence of the third site for Fedwire and it becomes significant only when the site failure rate is assumed to be larger than 2 failures per year. Since such high values are considered unrealistic, it is not clear whether the advantage of the Fedwire configuration over ACH is sufficiently high. This issue will be examined in more detail in the next section, along with analysis of other alternative configurations.

3.3 Analysis of Results: Existing System - Unreliable Communications.

In this subsection we consider a more realistic model of system operation, by assuming that the communication lines are unreliable. As with the processors, we assume that the operating and repair times for the communications lines follow exponential distributions independent of the corresponding quantities for the mainframes and the sites. A Markov process formulation is still possible under this assumption, however the resulting model is much more complicated than the first. The reason is that the information on which sites, mainframes and communications connections are working at some time is sufficient to determine the probability distribution of the future events, but not the operating condition of Fedwire or ACH. To see this, consider the following example. If it is known that site CC3 is the only site working

and the connections between CC1 and CC3 along with CC2 and CC3 are down, it cannot be determined whether Fedwire is operational. Given the current system configuration, to determine this event, it must also be known whether the most recent connection failure occurred before or after the most recent site failure. However, in order to capture this information the state description must be enhanced to a degree prohibitive for exact computational analysis.

To avoid enlargement of the state space we adopt the following simplification. We assume that the system is inoperable whenever two communication lines are inoperable. Under this assumption the problem size is greatly reduced because communication line failures can now be considered completely independently of site and mainframe failures. The analytical results of the simplified model provide an upper bound for the expected downtime and a lower bound for the availability of Fedwire with unreliable communications and the existing configuration. A similar argument also holds for ACH. Figures 3 and 4 are the analogs of figures 1 and 2 for the case of communication lines failure rate equal to 1 expected failure every 5 years.

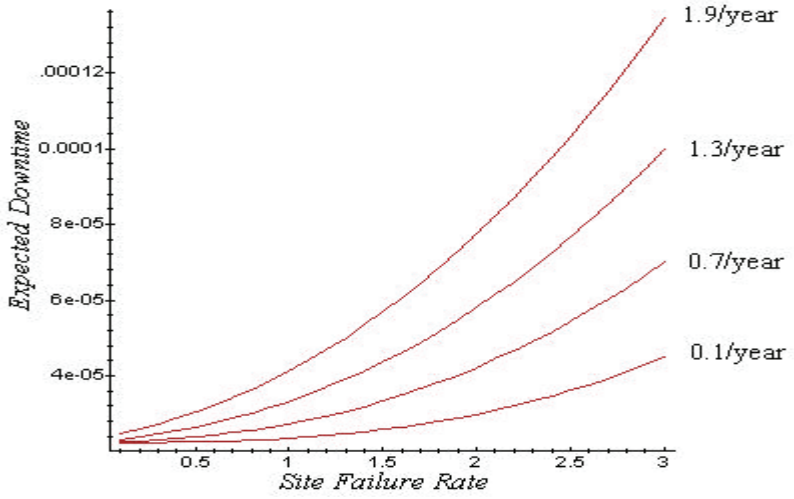

Figure 3: Expected Downtime of Fedwire (Unreliable Communications)

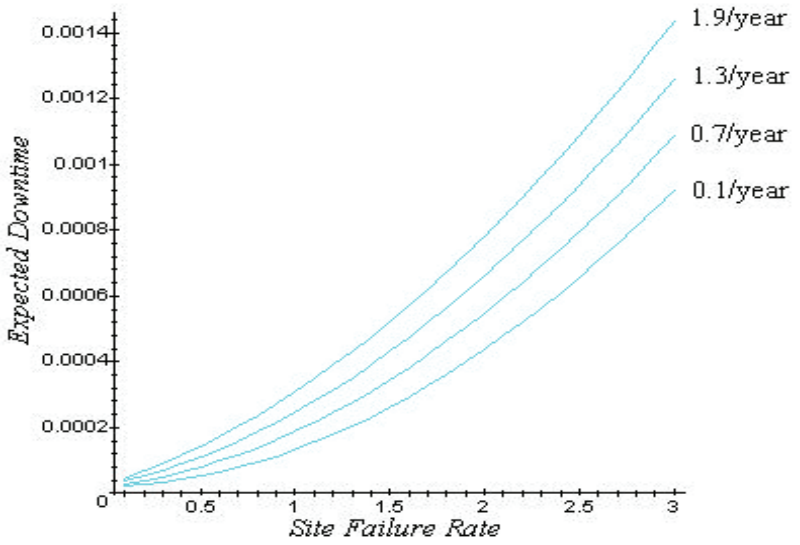

Figure 4: Expected Downtime of ACH (Unreliable Communications)

Figure 5 displays the difference between expected downtime of Fedwire with communication line failures and the expected downtime of Fedwire without communication line failures as computed by the model, for a value of the machine failure rate λ_M equal to 0.4. The horizontal line depicts the expected downtime of the communication lines, which does not depend on the site failure rate. Figure 6 presents the corresponding difference for ACH.

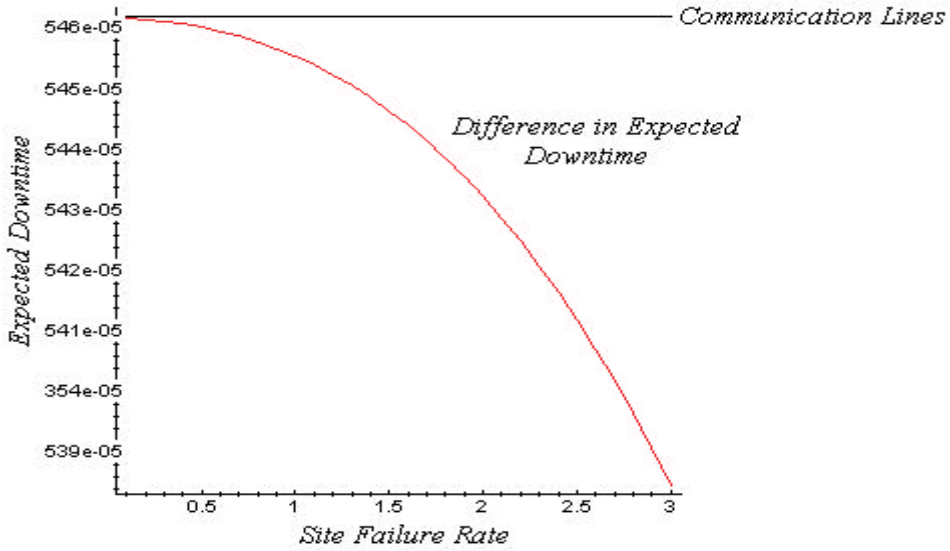

Figure 5: Difference of Expected Downtime for Fedwire with and without Communication Failures

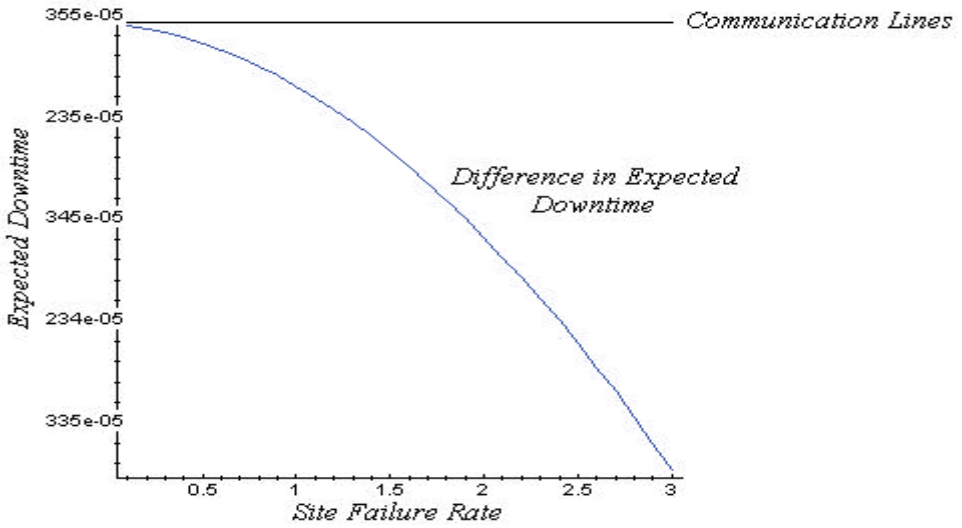

Figure 6: Difference of Expected Downtime for ACH with and without Communication Failures

A common observation in the above figures is that the difference in the expected downtime between the two models is decreasing and approaches zero as the site failure rate increases. This is expected, because as the site failures increase, the effect of the communication lines failures to the system downtime decreases.

4. Analysis of Alternative Configurations and Policies

The modeling approach developed in the previous section can be used to analyze the reliability of the system under hypothetical changes in the design and operation policies. In this section we consider two types of modifications. The first refers to alternative configurations that use a smaller number of processing sites and/or mainframes. The analysis is based on the observation made in the previous section that the difference in availability between Fedwire and ACH is small for small or moderate values of the site failure rate. This observation is consistent with the experience in the design of highly reliable processes. It is generally well known that the marginal reliability benefit from standby backup units is smaller at higher reliability levels. The second kind of modifications considered refers to using different

backup policies under the original system configuration. All the alternative configurations considered assumed completely reliable communication lines.

4.1 Alternative Configurations

The first alternative configuration is motivated by the comparison between Fedwire and ACH in the previous section. We consider a system in which the mainframes at site CC3 is eliminated, but communications can still be passed on to CC1 via CC3. Without considering the communication line failures, the expected downtime of Fedwire is presented in Figure 7.

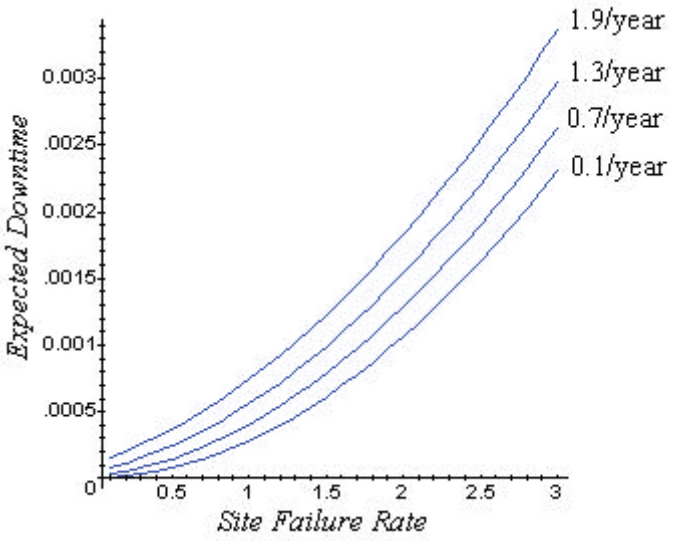

Figure 7: Expected Downtime for Fedwire of System without the CC3 mainframe

Although the expected downtime is still very low for this model, it is significantly higher than that of the original three-site system in Figure 1. This deterioration is expected partly because of the reduction in backup mainframe units. However the most important reason is that, due to the reduction in the number of sites from three to two, the site failures become the major cause of downtime.

The elimination of site CC3 resulted in an increase in the expected downtime of Fedwire from approximately 0.003% in Figure 1 to approximately 0.01% in Figure 7. An important reason for this increase is the reduction of backup sites for CMC from two to one. A simple way to correct the depletion of CMC in the modified model would be to add another backup on any of the mainframes in CC2. This may not be possible with the existing mainframe capacities, because of limited computer space. However, it is worthwhile to examine the implications of such a change, in order to estimate to what degree the benefit of CC3 is due to the CMC backup. The results are presented in Figure 8. As expected, the expected downtime is decreased significantly in comparison to that in Figure 7. This verifies the original conjecture that the CMC backup plays is the prominent reason for the importance of site CC3. It also follows from these figures that the machine failure rate is more critical in the two-site system without the additional CMC backup (Figure 7) than in the system with the backup (Figure 8).

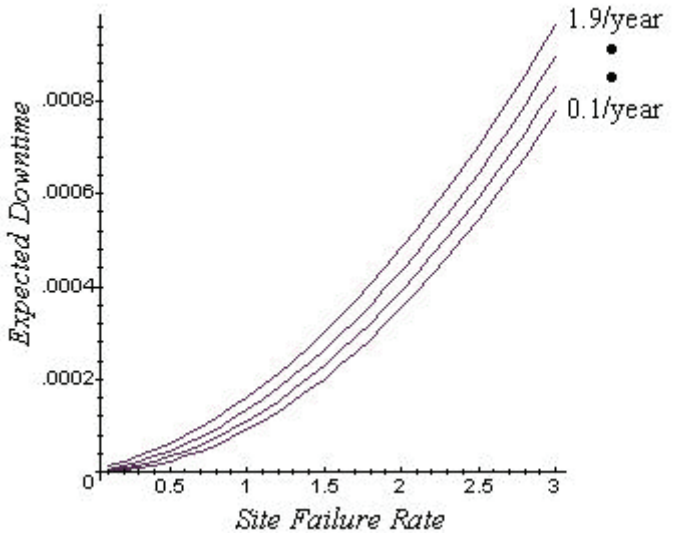

Figure 8: Expected Downtime for Fedwire of System without CC3 and two backups of CMC

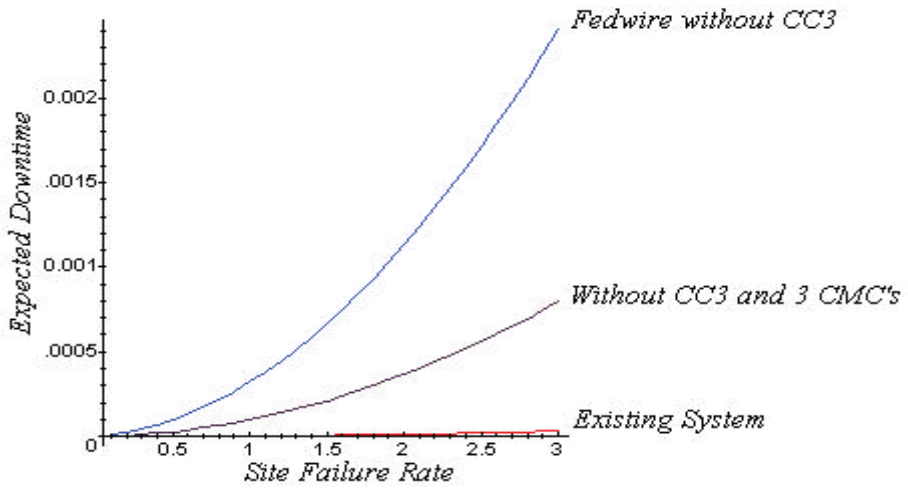

Figure 9: Comparison of the three configurations

In order to gain a better understanding of how the current system compares with the two alternative configurations considered in this section, Figure 9 presents the expected downtime of Fedwire for all three models, with common mainframe failure rate of 0.3 failures per year. All three different configurations offer very high reliability with the current system being the most reliable, as expected.

4.2 Alternative Recovery Policy

This subsection focuses on alternative recovery policies. Before proceeding with the analysis, it should be mentioned that recovery policies couldn't be changed without taking into account the system configuration. In designing a recovery policy two main factors are considered: computing resources and reliability. A backup policy may increase reliability but may cause certain mainframes to be overused, or it may be infeasible because of insufficient computing resources. Because in this study we have not considered computing capacity issues, the analyzed cases may not be entirely realistic under the current configuration. However they offer insight to the properties of the current system and can be considered as possible alternatives during planning for new resource acquisitions.

The policy change is related to the recovery of CMC, because, as was discussed previously, CMC constitutes a critical component of the payments system reliability. Under the current recovery policy, CMC may be transferred to another site although the mainframe currently operating CMC is working. If the policy were to transfer CMC only when the mainframe or site where it was currently being operated failed, the availability of Fedwire and ACH would increase. This is indicated by Figure 10 for Fedwire, with $\lambda_M = 0.4$ expected failures/year.

It should be noted that the analysis above is made under the assumption that all mainframes fail independently of each other. It is possible that if at one site a mainframe fails then the probability that a second mainframe will fail at the same site will increase. In this case whether CMC should be transferred or not would depend on the actual increase of the failure rate.

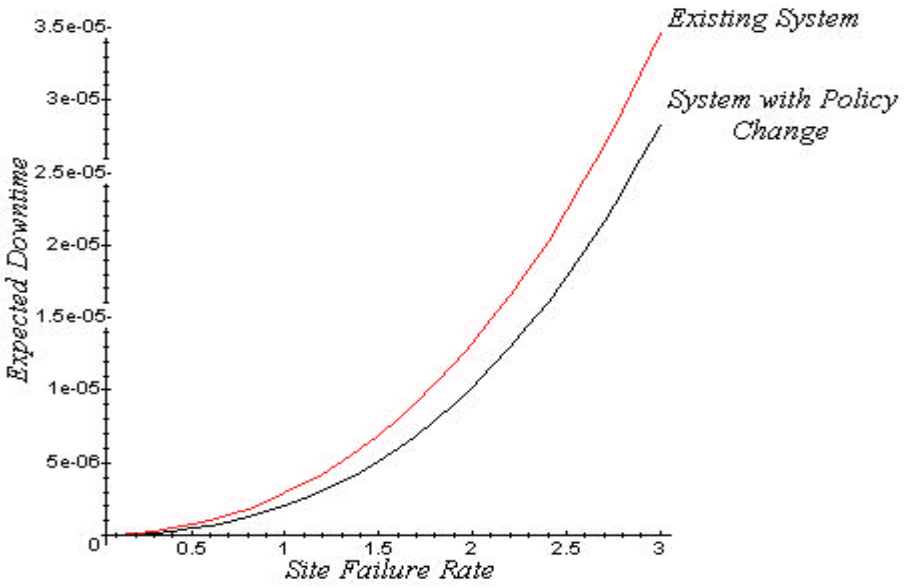

Figure 10: Expected Downtime of Fedwire for Current and New Policy

5. Conclusion

The reliability of the FRAS automation infrastructure was analyzed using models from the mathematical theory of random processes and applied probability. It was found that the system currently under operation displays a very high degree of reliability, because of extensive standby replication incorporated in the design. Because the costs of marginal reliability increases in highly reliable systems are typically great, it is possible for a slightly lower reliability level to be maintained with significantly lower standby replication. This, in turn, translates into substantial savings in investment and operating costs. Alternative configurations have been proposed in this spirit.

It must be mentioned that the compromise in reliability for the alternative systems and/or policies is very small. The analysis of the current Fedwire system under the assumption that one failure occurs every two years on average leads to typical values for expected downtime in the order of 1×10^{-6} (1 hour per 1,000,000 hours of operation). The alternative configuration where CC3 is completely eliminated but CMC is still backed up three times under the same failure rates leads to expected downtime in the order of 3×10^{-5} (3 hours per 100,000 hours of operation). One must ask whether such a difference in the availability of the system justifies the added cost of a third processing site.

6. Appendix: Mathematical Formulation

The results presented and discussed in Sections 3 and 4, regarding the reliability properties of the existing and alternative FRAS system configurations, were derived by modeling the operation of the entire system as a continuous time Markov process $\{X_t, t \geq 0\}$, where X_t denotes the system state at time instant *t*. We describe in detail the model used for the analysis of the current system assuming reliable communication connections, in Section 3 (Modeling Approach). The models for the other cases are similar.

The state in Section 3.1 (Modeling Approach) is described by a vector $x = (x_1, x_2, x_3)$, where x_i denotes the operating condition of processing site CCi, for *i*=1,2,3. Each parameter x_i takes values in a finite set, representing the possible states each processing site can be in. As discussed in 3.1, the state of a processing site is determined by the state of the mainframes comprising this site. In particular, given the current system configuration, as well as the standby/recovery policy in effect, the possible states for each site are presented below.

For site CC1, $x_1 \in \{0,1,2,3\}$, with the following state definitions:

state	*Description*
0	CC1 is in failure: This state is realized when the entire site is inoperable or two mainframes become inoperable at CC1.
1	CC1 is in Site Setup: This state is realized only after state 0. Once all mainframes have been repaired and the site is available for normal use this state represents a waiting period until it is brought back into use.
2	CC1 is Operational: This state is realized when all the mainframes and the site are in working order and being used.
3	CC1 is in Local Recovery: This state is realized when one of the two mainframes at CC1 is inoperable.

For site CC2, $x_2 \in \{0,\ldots,8\}$, where,

State	*Description*
0	CC2 is in failure: This state is realized when the entire site is inoperable or two mainframes become inoperable at CC1.
1	CC2 is in Site Setup: This state is realized only after state 0. Once all mainframes have been repaired and the site is available for normal use this state represents a waiting period until it is brought back into use.
2	CC2 is Operational: This state is realized when all the mainframes and the site are in working order and being used.
3	Primary mainframe failed
4	Primary mainframe under setup: This state is realized only after state 3 and represents the time after the unit has been repaired and before it is brought back into use.
5	Backup mainframe failed.
6	Backup mainframe under setup.
7	Disk storage unit failed.
8	Disk storage unit under setup.

For site CC3, $x_3 \in \{0,\ldots,6\}$, where,

state	*Description*
0	CC3 is in failure: This state is realized when the entire site is inoperable or two mainframes become inoperable at CC1.
1	CC3 is in Site Setup: This state is realized only after state 0. Once all mainframes have been repaired and the site is available for normal use this state represents a waiting period until it is brought back into use.
2	CC3 is Operational: This state is realized when all the mainframes and the site are in working order and being used.
3	Disk storage unit failed.
4	Disk storage unit under setup.
5	Primary mainframe failed.
6	Primary mainframe under setup.

From the above discussion it follows that the state space of the entire system can be represented by the product $S = \{0,\ldots,3\} \times \{0,\ldots,8\} \times \{0,\ldots,6\}$, and its cardinality is equal to $N = |S| = 252$. Because the state space is finite, it is possible to construct a one-one correspondence between the state vectors $x \in S$ and the integer numbers $i = 1,\ldots,N$. In the sequel a state will be denoted either by the vector or the integer number notation, as appropriate.

To describe the transitions between states, it is assumed that all times associated with mainframe or site failures, repairs, and setup are independent random variables following exponential distributions with appropriate parameters. In particular, the parameter for the distribution of a mainframe failure time (mainframe failure rate) is equal to λ_M, for all mainframes. In addition, the failure rate of catastrophic site failures is equal to λ_S, the repair rate of a mainframe equal to r_M, the repair rate of a failed site equal to r_S, the rate of the setup time distribution equal to s_S for site failures, and s_M for mainframes.

In order to analyze the probabilistic behavior of the model over time it is sufficient to describe the infinitesimal generator (transition rate matrix) Q, of the process, defined as an $N \times N$ matrix, where the general element Q_{ij} is equal to the rate of transition from state i to state j, for $i,j = 1,\ldots,N$. When an immediate transition from i to j is not possible, the corresponding rate is equal to zero. In addition, the diagonal elements of Q are defined as $Q_{ii} = -\sum_{j \neq i} Q_{ij}$, so that the row sums of Q are all equal to zero.

In the model described in Section 3.1 the infinitesimal generator Q can be constructed by considering, for each possible state, all events (failures, repairs, or setup completion) that can change the state, as well as the corresponding transition rates. For example, if the state of the system were x = (2,0,0) (all sites in failure except CC1 which is fully operational) we would have the following possible transitions with there appropriate rates:

States	Transition Rates
(3,0,0)	$2*\lambda_M$ (because failure of one of the two mainframes at CC1 would cause this state transformation.
(0,0,0)	λ_S (A site failure at CC1)
(2,1,0)	r_S (CC2 is repaired)
(2,0,1)	r_S (CC3 is repaired)

In the model described above all states are positive recurrent (ergodicity property). This is so because all failed mainframes and sites start being repaired immediately and all operating mainframes and sites are always subject to failure. Therefore, there is always a sequence of transitions that can, with positive probability, lead the system from any state to any other. In the finite state space case, this is a sufficient condition for positive recurrence.

The long-run behavior of the system can be described by the steady-state probability vector $\pi = (\pi_1, \pi_2 \ldots \pi_N)$, where π_i denotes the expected proportion of time that the system is in state *i*. Under the ergodicity property, the steady-state vector can be obtained as the unique solution to the following system of linear equations (steady-state equations):

$$\pi^T Q = 0, \text{ and } \sum_{i=1}^{N} \pi_i = 1.$$

After vector π is computed, the availability of Fedwire and ACH can be calculated as the sum of the steady-state probabilities for all states in which Fedwire and ACH, respectively, are operational. Alternatively, let $S_F=\{i=1,\ldots,N$: Fedwire is not operational in state $i\}$, and $S_A=\{i=1,\ldots,N$: ACH is not operational in state $i\}$. Then, the expected downtime of Fedwire and ACH, presented and discussed in Section 3 (Modeling Approach), is given by $d_F = \sum_{i \in S_F} \pi_i$ and $d_A = \sum_{i \in S_A} \pi_i$, respectively.

To determine the subsets S_F and S_A of the state space, one must consider which mainframes and sites are operational in each state, as well as the recovery/backup contingency plans in effect. Specifically for the model considered in section 3 (Modeling Approach), S_F consists of 48 states and S_A *of* 104.

The results presented in Section 3.2 were derived by solving the above steady state equations repeatedly for various values of the machine and site failure rate parameters. The results in Sections 3.3 and 4 were derived by solving the steady state equations of analogous continuous time Markov process models, according to the specific system configuration and recovery policy considered in each case.

References

Bauer, P. W. and Ferrier, G. D. (1996), "Scale Economies, Cost Efficiencies and Technological Change in Federal Reserve Payments", Federal Reserve Financial Services Working Paper Series, No. 01-96.

Gertsbakh, I. B. (1989), "*Statistical Reliability Theory*", Marcel Dekker, New York.

Karlin, S. and Taylor, H. (1981), "*A First Course in Stochastic Processes*", Academic Press, Boston.

Kao, E. P. C. (1997), "*An Introduction to Stochastic Processes*", Duxbury Press, Belmont, CA.

Keilson, J. (1987), "Robustness and Exponentiality in Redundant Repairable Systems", *Annals Operations Research*, **9,** pp. 439-447.

United States General Accounting Office, 1997, Payments, Clearance, and Settlement: A Guide to the Systems, Risks, and Issues (June 17), GAO/GGD-97-73.