УДК 519.687.1 : 519.687.4

Специализированная система электронного документооборота

О.И. Глебов

Предметом исследования является технология удаленного представления электронной отчетности по открытым каналам связи с использованием средств криптографической защиты информации (СКЗИ). В настоящей статье рассматриваются вопросы создания, функционирования и развития системы сдачи отчётности в электронном виде по телекоммуникационным каналам связи. В ходе работы установлены цели создания, назначение и основные задачи системы сдачи электронной отчетности по телекоммуникационным каналам связи с использованием средств криптографической защиты информации. Разработана схема организации электронного документооборота, выделены основные подсистемы.

Введение.

В работе рассматриваются вопросы создания, функционирования и развития системы сдачи отчётности в электронном виде по телекоммуникационным каналам связи.

В настоящее время достаточно широкое распространение получили компьютерные технологии, позволяющие организовывать подготовку документов бухгалтерской и налоговой отчетности в электронном виде. К ним относятся: 1С, Баланс 2W и др.

Электронная отчетность по телекоммуникационным каналам связи - это форма электронного документооборота по электронным сетям передачи информации, в частности – сети Интернет, с использованием:

- электронной цифровой подписи (ЭЦП);
- средств криптографической защиты информации (СКЗИ).

Электронный вид документооборота - это насущная необходимость, т.к. только электронная форма позволяет оперативно вводить, обрабатывать и анализировать документы.

В связи с этим возникла задача разработки системы, позволяющей сдавать отчетность в электронном виде по телекоммуникационным каналам связи.

1. Постановка задачи

Целью работы является создание такой системы для обеспечения юридически значимого документооборота, которая должна:

1. Обеспечить идентификацию отправителя и достоверность документа с помощью электронной цифровой подписи (далее ЭЦП).

2. Гарантировать сохранность и безопасность передаваемого документа с помощью средства криптографической защиты информации (далее СКЗИ).
3. Контролировать номенклатуру и сроки предоставления документов.

Требуется:

1. Разработать модель специализированной системы электронного документооборота.
2. Дать функциональное описание работы системы.
3. Выбрать и оценить показатели качества.
4. Разработать алгоритмы обработки заявок.
5. Исследовать входящие потоки.
6. Исследовать влияние входных потоков на характеристики системы.
7. Рассмотреть различные варианты работы системы.
8. Дать предложения по совершенствованию системы.

Работа ведется в рамках проекта с МКНТ (Московский комитет по науке и технологиям).

2. Схемы работы системы

В работе рассматриваются две схемы:

1. Передача документов с использованием электронной почты.
2. Подготовка и передача документов с использованием технологии “тонкого клиента” на основе web-браузера.

В случае использования электронной почты это:

- Передача электронного почтового сообщения, содержащего отчетность.
- Своевременная обработка его на стороне сервера.

В случае использования веб-браузера:

- Подготовка и заполнение отчетности через интернет
- Сдача отчетности.
- Своевременная обработка документов на стороне сервера.

Принцип работы системы основан на обмене электронными документами, заверенными ЭЦП.

2.1. Схема документооборота

Схема документооборота на примере сдачи налоговой отчетности представлена на рисунке 1.

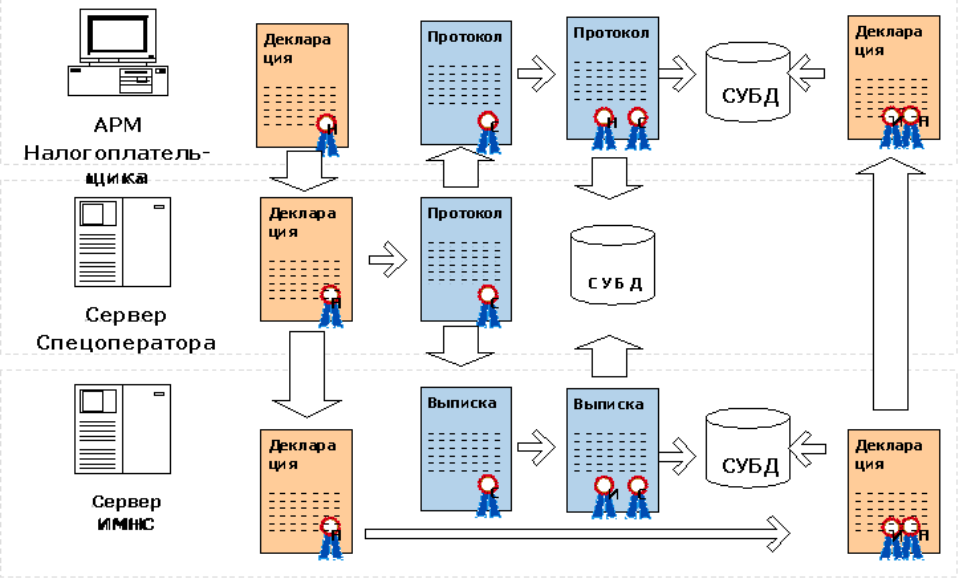


Рис. 1. Схема документооборота на примере сдачи налоговой отчетности

2.2. Схема организации связи

Общая схема организация связи, представленная на рис. 2, описывает использование телекоммуникационных средств, используемых в системе предоставления налоговой и бухгалтерской отчетности в электронном виде в региональном сегменте г. Москвы.

Представленная схема соответствует "Методическим рекомендациям об организации и функционировании системы предоставления налоговой и бухгалтерской отчетности в электронном виде по телекоммуникационным каналам связи" (утверждены приказом МНС РФ №БГ-3-32/705@ от 10 декабря 2002 г.).

На схеме обозначены следующие уровни:

- Уровень налогоплательщика;
- Уровень оператора связи;
- Уровень УМНС РФ по г. Москве (уровень ЦОД в ИМНС №39);
- Уровень ИМНС.

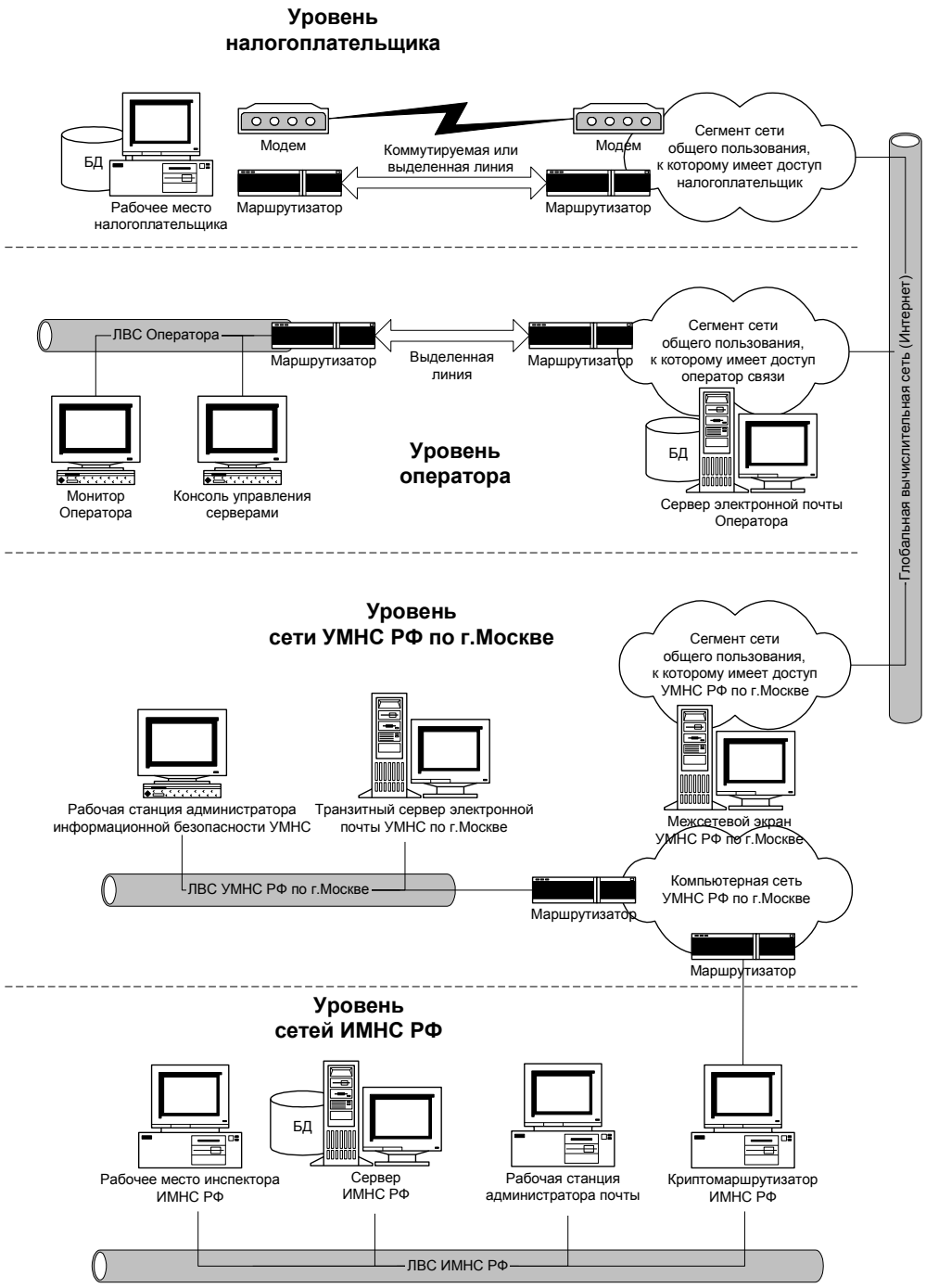


Рис. 2. Общая схема организация связи

3. Криптосистема

В начале 2002 года в России вышел закон об электронной цифровой подписи, который послужил толчком и основой для создания юридически значимого электронного документооборота.

В работе используется криптосистема RSA с открытым и секретным ключом.

3.1. Криптосистема RSA с открытым ключом

Криптосистемы с открытым ключом позволяют обмениваться секретными сообщениями по открытому каналу, не договариваясь заранее о ключе шифра; даже перехватив весь разговор от начала до конца, “противник” не узнает секретного сообщения. Кроме того, эти же методы позволяют добавлять к сообщению "цифровую подпись", удостоверяющую, что сообщение не фальсифицировано “врагами”. Проверить аутентичность подписи легко, а подделать её крайне трудно. Подобные методы находят широкое применение в банках, при подписывании контрактов, при денежных переводах и т. п.

Криптосистема RSA основана на таком обстоятельстве: в настоящее время известны эффективные алгоритмы поиска больших простых чисел, но не известно сколько-нибудь приемлемого по времени работы алгоритма разложения произведения двух больших простых чисел на множители.

3.1. Шифрование с открытым ключом

При использовании таких систем каждый участник переговоров имеет открытый ключ (public key) и секретный ключ (secret key). В системе RSA ключ состоит из двух целых чисел. Участников переговоров может быть несколько, но для примера мы будем говорить о переговорах участника A и участника B. Их открытые ключи мы будем обозначать $\mathbf{P}_\mathrm{A}$ и $\mathrm{P_B}$, а секретные - $\mathbf{S}_\mathrm{A}$ и $\mathbf{S}_\mathrm{B}$.

Каждый участник сам создаёт два своих ключа. Секретный ключ он хранит в тайне, а открытый сообщает остальным участникам (и вообще всем желающим, например, через газеты или Internet; открытые ключи можно публиковать в специальных справочниках и т. п.).

Обозначим через D множество всех возможных сообщений (например, это может быть множество всех битовых строк). Потребуем, чтобы каждый ключ задавал перестановку множества D, и через $\mathrm{P_A}()$ и $\mathrm{S_A}()$ будем обозначать перестановки, соответствующие ключам A. Мы считаем, что каждая из перестановок $\mathrm{P_A}()$ и $\mathrm{S_A}()$ может быть быстро вычислена, если только известен соответствующий ключ.

Мы хотим, чтобы ключи одного участника задавали взаимно обратные перестановки, т. е. чтобы

$$M = S_A(P_A(M))$$

(3.1.1.)

и

$$M = P_A(S_A(M)) \qquad (3.1.2.)$$

было выполнено для любого сообщения M C D

Самое главное - чтобы никто, кроме A, не мог вычислять функцию $S_A()$ за разумное время; именно на этом основаны все полезные свойства криптосистемы, перечисленные выше. Потому-то A и держит значение S_A в секрете: если кто-либо узнает его секретный ключ, он сможет расшифровывать адресованные A сообщения, подделывать его подпись или подменять сообщения, которые A отправляет от своего имени. Главная трудность при разработке криптосистем состоит в том, чтобы придумать функцию $S_A()$, для которой трудно было бы найти быстрый способ вычисления, даже зная такой способ для обратной функции $P_A()$.

3.2. Криптосистема RSA

Чтобы построить пару ключей для криптосистемы RSA (RSA cryptosystem), надо сделать следующее.

- Взять два больших простых числа p и q (скажем, около 100 десятичных цифр в каждом).
- Вычислить $n = pq$.
- Взять небольшое нечётное число e, взаимно простое с $j(n)$. Получаем $j(n) = (p-1)(q-1)$.
- Вычислить $d = e^{-1} \bmod j(n)$ (По следствию 33.26 (см. [1]) d существует и определено однозначно по модулю $j(n)$.)
- Составить пару $P = (e, n)$ - открытый RSA-ключ (RSA public key).
- Составить пару $S = (d, n)$ - секретный RSA-ключ (RSA secret key).

Множеством D всех возможных сообщений для этой криптосистемы является Z_n Открытому ключу $P = (e, n)$ соответствует преобразование

$$P(M) = M^e \bmod n \qquad (3.1.3.)$$

а секретному ключу $S = (d, n)$ - преобразование

$$S(C) = C^d \bmod n \qquad (3.1.4.)$$

Как уже говорилось, эти преобразования можно использовать и для шифрования, и для электронных подписей.

Если считать, что числа d и n имеют порядка b битов, а число e имеет $O(1)$ битов, то преобразование P потребует O(1) умножений по модулю n ($O(b^2)$ битовых операций), а преобразование S потребует $O(b)$ умножений ($O(b^3)$ битовых операций) (разумеется, при известном ключе).

Теорема 1 (корректность системы RSA) Формулы (3.1.3.) и (3.1.4.) задают взаимно обратные перестановки множества Zn

Доказательство. Очевидно,

$$P(S(M)) = S(P(M)) = M^{ed} \bmod n$$

для всякого M C Zn Мы знаем, что e и d взаимно обратны по модулю $j(n)$, т. е.

$$ed = 1 + k(p-1)(q-1)$$

для некоторого целого k. Если M =/= 0 (mod p), то по малой теореме Ферма (см. [1] теорема 33.31) имеем

$$M^{ed} \equiv M(M^{p-1})^{k(q-1)} \equiv M * 1^{k(q-1)} \equiv M$$

по модулю p. Равенство $M^{ed} \equiv M \pmod p$ выполнено, конечно, и при $M \equiv M \pmod p$, так что оно верно для всех M. По тем же причинам $M^{ed} \equiv M \pmod q$, и потому $M^{ed} \equiv M \pmod n$ при всяком M.

Надёжность криптосистемы RSA основывается на трудности задачи разложения составных чисел на множители: если «враг» разложит (открыто опубликованное) число n на множители p и q, он сможет найти d тем же способом, что и создатель ключа. Таким образом, если задача разложения на множители может быть решена быстро (каким-то пока неизвестным нам алгоритмом), то "взломать" криптосистему RSA легко. Обратное утверждение, показывающее, что если задача разложения на множители сложна, то взломать систему RSA трудно, не доказано - однако за время существования этой системы никакого иного способа её взломать обнаружено не было.

Разложение чисел на множители - дело непростое, быстрого алгоритма для этого неизвестно; известные ныне методы не позволяют разложить на множители произведение двух 100-значных простых чисел за разумное время - для этого нужны какие-то новые идеи и методы (если это вообще возможно).

Конечно, надёжность системы RSA зависит от размера простых чисел, поскольку небольшие числа легко разложить на множители. Поэтому надо уметь искать большие простые числа.

На практике для ускорения вычислений криптосистему RSA часто используют вместе с какой-то традиционной системой шифрования, в которой ключ необходимо хранить в секрете. Выбрав такую систему, мы используем для шифрования её - а система RSA используется только для передачи секретного ключа, который может быть значительно короче самого сообщения. Сам этот ключ может выбираться, например, случайно и только на один раз.

Похожий подход применяется для ускорения работы с цифровыми подписями. Система RSA используется при этом в паре с так называемой односторонней хеш-функцией (one-way hash function). Такая функция отображает каждое сообщение M в достаточно короткое сообщение $h(M)$ (например, 128-битовую строку), при этом $h(M)$ легко вычислить по M, но

не удаётся найти два разных сообщения M и M', для которых $h(M) = h(M')$ (хотя таких пар много по принципу Дирихле).

Образ $h(M)$ сообщения M можно сравнить с "отпечатком пальца" (fingerprint) сообщения M. A, желая подписать своё сообщение M, вычисляет $h(M)$, а затем шифрует $h(M)$ своим секретным RSA-ключом. Затем A посылает B пару $(M, S_A(h(M))$. B удостоверяется в подлинности подписи, проверив, что $P_A(S_A(h(M))) = h(M)$. Конечно, можно фальсифицировать текст сообщения, найдя другое сообщение M', для которого $h(M) = h(M')$, но это (по предположению) сложно.

Конечно, при использовании открытых ключей надо ещё убедиться, что сами ключи не были подменены. Предположим, что имеется некоторый "нотариус" (trusted authority), честность которого вне подозрений и открытый ключ которого все знают (и в его правильности не сомневаются). В книге [2] используется термин "центр доверия". Нотариус может выдавать известным ему людям справки (сертификаты, certificates) о том, что их открытый ключ такой-то, подписывая эти справки собственной цифровой подписью. (Приходящие должны сообщить нотариусу свой открытый ключ.) Подлинность сертификата может быть проверена каждым, кому известен открытый ключ нотариуса; любой зарегистрированный у нотариуса участник переговоров может прилагать к своим сообщениям выданный нотариусом сертификат.

В работе используется СКЗИ (средство криптографической защиты информации) КриптоПро CSP, которое является средством криптографической защиты информации, разработанным ООО "Крипто-Про". СКЗИ КриптоПро CSP реализует российские криптографические алгоритмы и разработана в соответствии с криптографическим интерфейсом фирмы Microsoft - Cryptographic Service Provider (CSP).

4. Схема информационного обмена

На рис. 3 проиллюстрирована процедура электронного документооборота на примере сдачи налоговой отчетности.

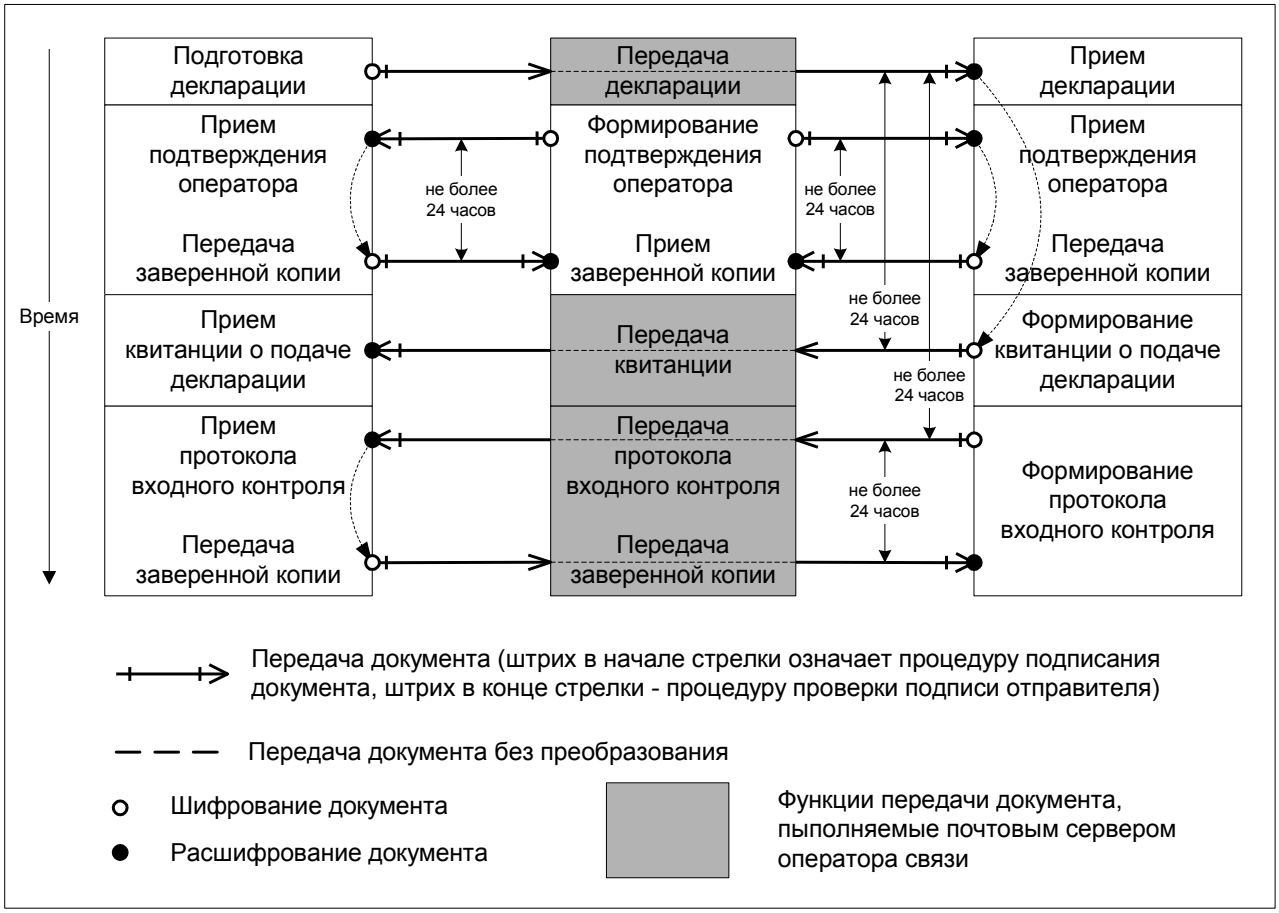


Рис. 3. Электронный документооборот на примере сдачи налоговой отчетности

Как видно из рисунка, в процессе обработки документов очень важным становится фактор времени. Система электронного обмена должна быть устойчивой к пиковым нагрузкам. Все документы на сервере должны быть обработаны и отправлены в срок.

В связи с этим, особое внимание следует уделить повышению производительности сервера.

4.1. Экспертная оценка

На следующих схемах представлена экспертная оценка количества деклараций, сдаваемых по каналам связи, в зависимости от количества подключенных участников информационного обмена.

Экспертная оценка количества деклараций, сдаваемых по каналам связи, в зависимости от количества подключенных участников информационного обмена (налогоплательщиков).

4.1.1. Количество сдаваемых деклараций

Количество участников: 2000

Табл. 1

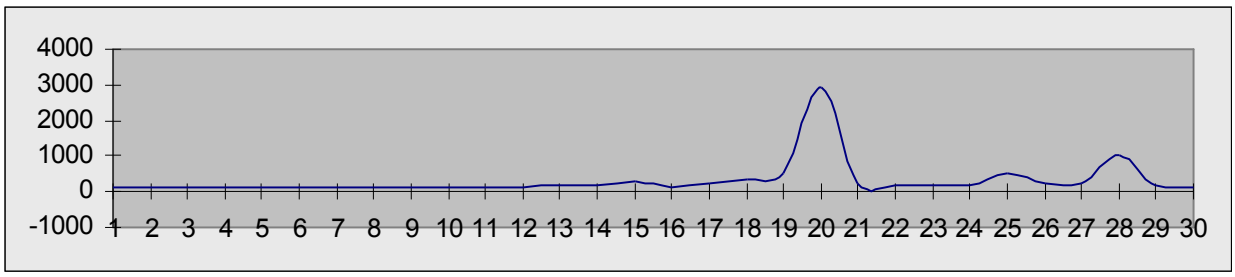
Сдает декларации	срок сдачи	Количество, сдающих		Процент деклараций от сдающих		Средний размер (Кб)	Всего (Кб)	
		Ежеме-сячно	Ежеквар-тально	Ежеме-сячно	Ежеквар-тально		Еже-месячно	Ежеквар-тально
НДС	20	1000	2000	50%	100%	8	8000	16000
авансы по ЕСН	20	1000	2000	50%	100%	8	8000	16000
авансы в ПФ	20	1000	2000	50%	100%	8	8000	16000
налог на прибыль	28	1000	2000	50%	100%	8	8000	16000
Бухгалтерский .баланс (Ф 1)	30		2000	0%	100%	8	0	16000
Отчет о прибылях и убытках (Ф 2)	30		2000	0%	100%	8	0	16000
Налог на имущество	30		2000	0%	100%	8	0	16000
Прочие документы по сроку 15 (акцизы)	15	200	200	10%	10%	8	1600	1600
Прочие документы по сроку 20 (игорный бизнес)	20	200	200	10%	10%	8	1600	1600
Прочие документы по сроку 25 (акцизы и упрощенка)	25	400	400	20%	20%	8	3200	3200
Итого документов по сроку 15 число		200	200			8	1600	1600
Итого документов по сроку 20		3200	6200			8	25600	49600
Итого документов по сроку 25		400	400			8	3200	3200
Итого документов по сроку 28		1000	2000			8	8000	16000
Итого документов по сроку 30		0	6000			8	0	48000
Итого документов		4800	14800			8	38400	118400

4.1.2. Количество деклараций, сдаваемых ежемесячно

Табл. 2

Количество деклараций, сдаваемых ежемесячно	1-ое*	2-ое	3	4	5	6	7	8	9	10	11	12	13	14	15
	144	144	144	144	144	144	144	144	144	144	144	148	158	168	318
	16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
	148	208	368	528	2928	208	152	172	192	512	222	264	1014	164	144

* - число месяца, следующего за отчетным периодом

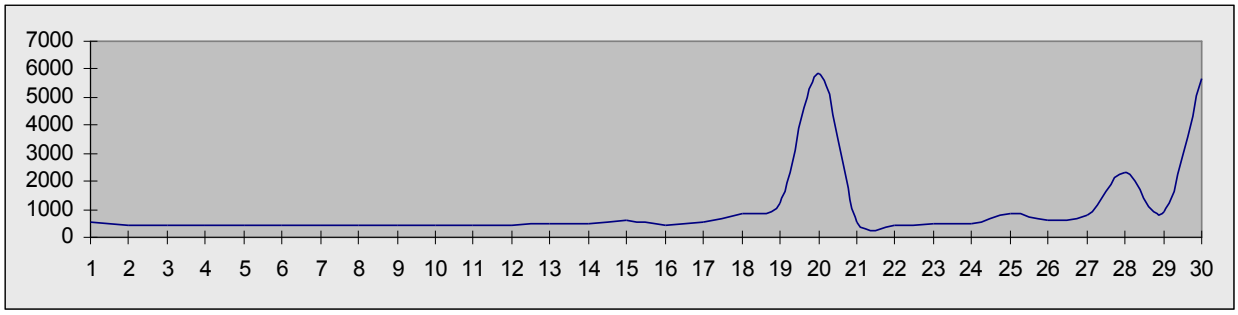


Количество деклараций

4.1.3. Количество деклараций, сдаваемых ежеквартально

Табл. 3

Количество деклараций, сдаваемых ежеквартально	**1-ое**	**2-ое**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
	558	444	444	444	444	444	444	444	444	444	444	448	458	468	618
	16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
	448	568	878	1188	5838	568	452	472	492	832	592	804	2304	904	5664

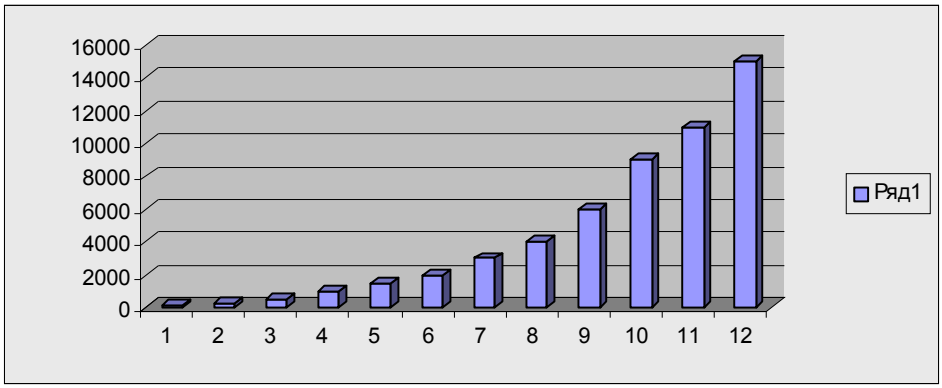


Количество деклараций

4.1.4. План подключения количества налогоплательщиков к системе сдачи отчетности по каналам связи

Табл. 4

Кварталы	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Количество налогоплательщиков	100	300	500	1000	1500	2000	3000	4000	6000	9000	11000	15000



План подключения

4.2. Производительность серверов

Как видно из графиков, существуют пиковые дни, когда подготавливается и сдается большое число деклараций. В такие дни нагрузка на сервер максимальна, что может привести к резкому снижению производительности сервера, и как следствие увеличению времени обработки документов.

Производительность сервера становится особенно важна в случае использования схемы сдачи через веб-браузер, когда в процессе подготовки налоговой декларации требуется мгновенный отклик сервера.

5. Системы распределения нагрузки.

5.1. Способы повышения быстродействия обработки информации

Существует несколько способов повышения быстродействия:

1. можно увеличить полосу пропускания,
2. установить высокопроизводительное сетевое оборудование,
3. разработать эффективные приложения,
4. оптимизировать и модернизировать программные и аппаратные компоненты сервера.

Еще один способ повышения производительности сервера состоит в том, чтобы увеличить количество серверов и размещать на них "зеркальные" копии материалов. Таким образом, можно распределить общую нагрузку по всем компонентам системы и сократить время возврата информации при выполнении сервером внутренних процедур обработки клиентских запросов.

Распределение, или выравнивание нагрузок, приходящихся на несколько серверов, позволяет избежать такой ситуации, когда передаваемые по сети Web пакеты лавиной обрушиваются на один сервер, в то время как другие простаивают без дела.

5.2. Система распределения нагрузки

Система распределения нагрузки из нескольких машин представлена на рис. 4

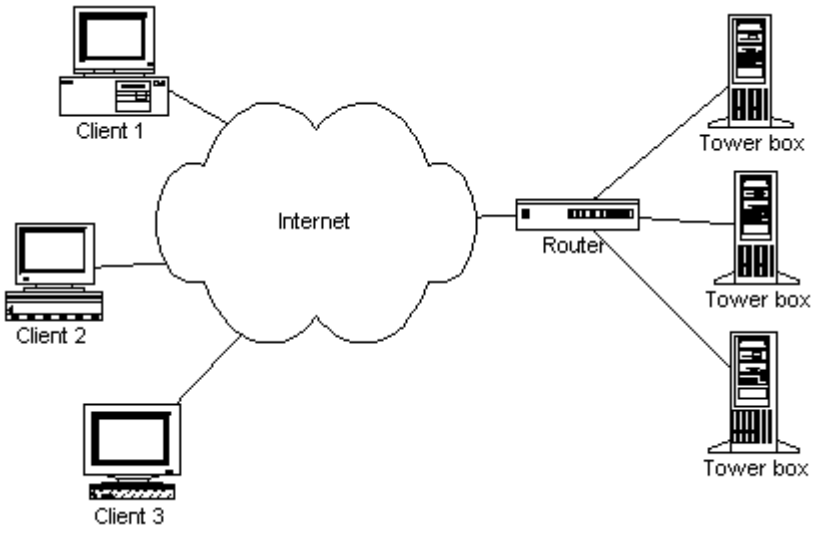


Рис. 4. Система распределения нагрузки из нескольких машин

Поступающие запросы распределяются между серверами. Сервер обрабатывает запрос, и при необходимости считывает данные из базы данных, или пишет в нее. Если кокой-то из серверов выходит из строя или максимально загружен, то запросы распределяются по оставшимся серверам. Подобная конфигурация обеспечивает распределение нагрузки (load balancing), т.е. нагрузка по обработке запросов распределяется между несколькими серверами. Так же эта конфигурация обеспечивает отказоустойчивать системы (fault tolerance), поскольку, при выходе из строя одного из серверов сам web-сайт продолжает работать.

5.3. Схемы распределения нагрузки во многомашинной системе

Существует несколько распространенных подходов в организации распределения нагрузки:

- круговой DNS, когда для распределения нагрузки используется DNS-сервер;
- аппаратное распределение нагрузки, когда используется прибор, схожий по своим функциям с маршрутизатором;
- программное распределение нагрузки. Например, программа "TCP/IP Network Load Balancing" от Microsoft;
- смешанные схемы, когда используется комбинация аппаратных и программных средств.

5.4. Системы сдачи отчетности через интернет

Предполагаемая схема построения системы сдачи отчетности через интернет показана на рис. 5.

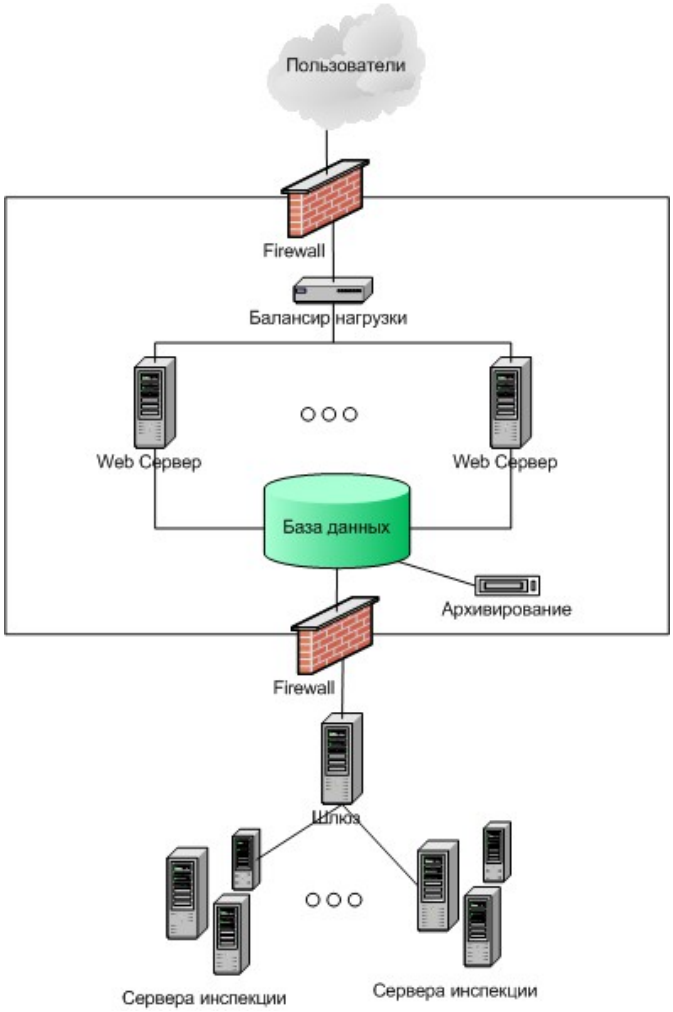


Рис. 5. Схема построения системы сдачи отчетности через интернет

Такая схема обладает следующими преимуществами:

- Масштабирование (Количество Web-серверов может наращиваться в зависимости от количества пользователей и загрузки существующих серверов).
- Производительность.
- Гибкость (Использование в системе нескольких серверов с разными характеристиками и настройка системы с учетом их различной производительности).
- Отказоустойчивость (В случае выхода из строя одного из серверов, остальные сервера принимают его нагрузку на себя).

5.4.1. Масштабируемость

Вне зависимости от конструкции в машину можно установить определенный максимум процессоров. Многие дешевые сервера позволяют установить лишь два процессора. Многие высококлассные дорогие сервера имеют предел в 16 процессоров. В системе распределения нагрузки может участвовать до 255 машин.

5.4.2. Производительность

Производительность сервера зависит от нескольких факторов: скорости работы процессора, памяти, системной шины, операционной системы и проч. В системе распределения нагрузки каждый сервер имеет свой собственный процессор, свой объем памяти, свою операционную систему и проч. В многопроцессорных же системах большинство ресурсов находятся в общем пользовании. Любой из перечисленных факторов может стать узким местом и тем самым ухудшить работу системы в целом, даже несмотря на добавление процессоров. В системе же распределения нагрузки прирост мощности линейно зависит от количества машин.

5.4.3. Гибкость

Во многопроцессорных машинах работает один тип операционной системы, один тип процессора, один тип памяти и так далее. Используя «спарку» серверов, есть возможность «смешивать» машины с различными операционными системами, процессорами и объемами памяти. Если потребуется, в систему распределения нагрузки можно добавить и многопроцессорную машину. Различным машинам в системе распределения нагрузки можно присвоить свои весовые коэффициенты, так что на более быстрые машины ляжет большая нагрузка, чем на более медленные.

6. Выводы

В данный момент реализован полный цикл сдачи отчетности со стороны налогоплательщика. В процессе работы получены следующие результаты:

1. Сформированы цели создания, назначение и главная задача системы.
2. Установлены требования к системе, основанные на следующих документах:
   - "Порядок предоставления налоговой декларации в электронном виде по телекоммуникационным каналам связи" утвержденным приказом МНС от 02 апреля 2002 №БГ-3-32/169 (далее – Порядок);
   - "Методические рекомендации об организации и функционировании системы предоставления налоговой и бухгалтерской отчетности в электронном виде по телекоммуникационным каналам связи" утвержденные Приказом МНС России от 10 декабря 2002 года №БГ-3-32/705@ (далее – Методические рекомендации).
3. Определена схема организации связи, включающая в себя:
   - Схему организации обмена по электронной почте;
   - Общую схему организации связи.
4. Определены требования к системе, включающие в себя:

- Требования к программному обеспечению участников документооборота;
- Требования к защите информации;
- Требования к надежности;
- Требования к стандартизации и унификации.

5. Развернут стенд для тестирования и отладки электронного документооборота.
6. Разработан макет приложения подготовки и сдачи налоговых деклараций по каналам связи.
7. В рамках веб-ориентированной версии был разработан макет программного обеспечения, позволяющий заполнять и сдавать налоговые декларации, используя веб-браузер.
8. В рамках почтовой (E-mail) версии были реализованы общие модули системы, включающие модуль почтового клиента и модуль СКЗИ:
   - Модуль почтового клиента выполняет функции формирования, отправки, получения и разбора почтовых сообщений.
   - Модуль СКЗИ позволяет формировать и проверять электронную цифровую подпись (ЭЦП), шифровать и расшифровывать данные. Модуль СКЗИ использует КриптоПро CSP.
   - Реализован упрощенный интерфейс к первым двум модулям, выполняющий основные функции отправки и получения налоговых документов по телекоммуникационным каналам связи.

Список литературы

1. Кормен Т., Лейзерсон Ч. и Ривест Р. Алгоритмы. Построение и анализ. - М: МЦНМО, 1999г. – 960 стр.

2. Введение в криптографию под ред. В. В. Ященко, М.: МЦНМО, 2000г. – 288 стр.

3. Вентцель А. Д. Курс теории случайных процессов. - М: Наука, 1975 г. - 320 стр.

4. Вентцель Е. С. Исследование операций. Задачи, принципы, методология. Учебное пособие для вузов. - М: Дрофа, 2004 г. - 208 стр.

5. "Методические рекомендации об организации и функционировании системы предоставления налоговой и бухгалтерской отчетности в электронном виде по телекоммуникационным каналам связи" (утверждены приказом МНС РФ №БГ-3-32/705@ от 10 декабря 2002 г.).

6. Тао Чжоу. Системы балансировки нагрузки Web-серверов // Windows 2000 Magazine. – 2000, №3.

- http://www.citforum.ru/internet/webservers/websbal.shtml

7. Качанов Александр. Схемы балансировки нагрузки для web-серверов.

- http://www.webmascon.com/topics/technologies/4a.asp

Глебов Олег Игоревич, аспирант кафедры математической кибернетики Московского авиационного института (государственного технического университета);
Телефон: 211-3324, e-mail: glebov@oviont.ru