

Р. Г. Шыхалиев, канд. техн. наук, вед. науч. сотр., ramiz@science.az
Институт Информационных Технологий НАНА, г. Баку

Повышение эффективности мониторинга компьютерных сетей на основе оптимизации системы поллинга

Предлагается методология повышения эффективности мониторинга компьютерных сетей (КС), которая позволяет минимизировать время, затрачиваемое на мониторинг. Основной целью при этом является оптимизация мониторинга сетевых узлов КС при заданных сетевых ресурсах. Для достижения этой цели предлагается использовать модель системы поллинга, оптимизируя которую можно достичь оптимизации мониторинга КС.

Ключевые слова: компьютерные сети, сетевой мониторинг, поллинг сетевых узлов, SNMP-поллинг, системы поллинга, дисциплина поллинга, среднее время ожидания

Введение

Непрерывно растущий масштаб и сложность современных компьютерных сетей (КС) увеличивают потребность в их автоматическом управлении. При этом время, затрачиваемое на принятие решений по управлению сетью, имеет существенное значение, особенно при нарушении безопасности сети и возникновении аварийных ситуаций в критических системах.

Ключевым компонентом автоматического управления КС является мониторинг в режиме реального времени, одной из задач которого является измерение характеристик сетевых узлов и каналов связи, позволяющих контролировать основные параметры сети. При этом для получения значений характеристик сетевых узлов и каналов связи может быть использован непосредственный поллинг узлов КС в режиме реального времени. Исходя из этого, мониторинг КС может быть смоделирован как система поллинга [1], что позволит в режиме реального времени непрерывно обновлять данные мониторинга и обеспечивать их актуальность.

Как правило, мониторинг КС осуществляется только в отношении основных (или критических) сетевых узлов [2], к которым относятся серверы, маршрутизаторы, коммутаторы, хабы и межсетевые экраны. Кроме основных сетевых узлов, при необходимости также может быть осуществлен мониторинг и других сетевых узлов, например рабочих станций, терминалов и т. д. Конечно, самой лучшей стратегией мониторинга КС был бы постоянный опрос всех ее узлов в реальном времени в целях получения всеобъемлющей информации о состоянии КС. Однако вследствие распространения задержки при поллинге монитором узлов сети и увеличения общего времени обработки всех запросов увеличивается время реагирования сетевой управляющей системы на возможные аварийные и аномальные ситуации, что особенно актуально в больших КС. В итоге снижается эффективность систем управле-

ния КС, т. е. эффективность систем управления сетями зависит не только от эффективности алгоритмов управления, но и от оптимальности мониторинга сети. Таким образом, оптимизация мониторинга КС, т. е. оптимизация поллинга узлов КС, является одной из главных задач по управлению КС. Следовательно, необходимо разработать новый подход к мониторингу КС для оптимизации его процесса. При этом оптимизация процесса мониторинга позволит уменьшить время реагирования сетевой управляющей системы на инциденты и аварийные ситуации. Исходя из этого, в данной работе для мониторинга КС предлагается использовать теорию систем поллинга.

В литературе имеются ряд работ, которые посвящены повышению эффективности мониторинга сетей путем оптимизации тех или иных аспектов поллинга. Например, в работе [3] авторы предлагают динамическую схему поллинга, которая позволяет оптимизировать мониторинг сети посредством оптимизации частоты опроса узлов сети. Предлагаемая схема поллинга позволяет определить частоту опроса для каждого узла сети, и для этого используются метод анализа иерархий и метод ветвей и границ. При этом основной целью является обеспечение эффективности управления сетью при достаточно высокой производительности сети.

В работе [4] для адаптивного мониторинга масштабируемых корпоративных систем авторы предлагают подход оптимизации плана поллинга, который позволяет изменять план мониторинга. Адаптивный мониторинг является перспективным методом автоматизации конфигурации сервера мониторинга корпоративных систем с динамической реконфигурацией. Однако генерация плана поллинга является NP-трудной задачей, и авторы предлагают приближенный алгоритм.

В работе [5] для повышения эффективности мониторинга и управления КС с помощью SNMP-поллинга авторы предлагают подход, основанный на эффективной передаче больших и динамически

изменяющихся MIB-баз между центром управления сетью и SNMP-агентами. Этот механизм, известный как GetModify, позволяет передавать данные, измененные только в интервалах между опросами SNMP-агентов монитором. Причем может быть вычислена скорость передачи данных, а также число пакетов и количество информации, переданной за единицу времени. В работе [6] авторами рассмотрена задача структурного синтеза систем мониторинга. Для этого предлагается система поддержки принятия решений по выбору структуры системы мониторинга и дисциплины поллинга узлов фрагмента КС.

Анализ работ, указанных выше, показал, что, несмотря на большой вклад в повышение эффективности мониторинга сетей, эти работы основываются на архитектурном решении задачи оптимизации тех или иных аспектов поллинга. Причем не рассматриваются характеристики самой системы поллинга, такие как интенсивность поступления заявок (данных мониторинга), среднее время ожидания заявок в очереди, время переключения сервера между очередями и т. д. Исходя из этого, в данной работе для повышения эффективности мониторинга КС предлагается оптимизировать саму систему поллинга, и поскольку одним из основных показателей оптимальности систем поллинга является минимальное среднее время ожидания заявок в каждой очереди, то для оптимизации поллинга узлов КС решено минимизировать среднее время ожидания заявок в очередях. На наш взгляд, это был бы единый подход для повышения эффективности мониторинга КС на основе системы поллинга.

1. Методы мониторинга компьютерных сетей

На сегодняшний день для сбора необходимой информации о состоянии КС в основном используется мониторинг. К задачам мониторинга относят: сбор значений характеристик каналов передачи и коммутирующего оборудования КС; выявление в КС аномальных ситуаций, а также узких мест; прогнозирование последствий изменений в топологии КС; мониторинг поведения пользователей КС и т. д. [7].

Существуют несколько способов мониторинга, которые позволяют минимизировать влияние процесса мониторинга на функционирование КС [8–10]. Один из способов основывается на ограничении трафика, генерируемого системой сетевого мониторинга. Метод в основном заключается в том, что сетевые администраторы в системе сетевого мониторинга могут устанавливать оптимальный предельный уровень для объема трафика, генерируемого ею. В результате трафик системы сетевого мониторинга будет минимизирован, и достоверность результатов повысится. Другой способ заключается в мониторинге не всех узлов КС, а только критических, таких как маршрутизаторы или серверы. Этот метод целесообразно применять в больших КС, где число

узлов измеряется тысячами и мониторинг всех узлов был бы бесполезным.

При мониторинге КС важен мониторинг не только сетевых устройств, но и приложений, выполняемых в них. Например, системному администратору необходимо быть уверенным, что операционные системы сетевых устройств (например, серверов, маршрутизаторов, коммутаторов и т. д.) сконфигурированы корректно и функционируют правильно. Кроме того, мониторинг приложений в основном используется для проверки текущего состояния сетевых узлов, например серверов. В зависимости от типа узла может быть осуществлен мониторинг специфических для данного узла характеристик. На основании этих характеристик можно определить состояние узла, а также факт превышения значения той или иной характеристики критического уровня. Вместе с тем при мониторинге КС должно быть принято решение о применении централизованного (т. е. мониторинг в одной точке) или распределенного мониторинга (т. е. одновременно в нескольких точках).

Нами были проанализированы некоторые существующие методы мониторинга [11–17], и SNMP (Simple Network Management Protocol) [18, 19] был выбран как основа для моделирования мониторинга КС.

SNMP является стандартным протоколом управления сетью и позволяет обеспечить interoperability систем управления сетью и широко использовать их для управления и мониторинга сетевых узлов в КС. SNMP определяет структуру иерархически организованной базы данных, которая хранится в сетевых узлах КС. В каждом узле содержится информация о предоставляемой услуге, таблице маршрутизации, а также различные статистические данные, отражающие их состояние, которые хранятся в базе данных MIB (Management Information Base) [20]. При этом должны быть измерены такие характеристики, которые описывают различные аспекты поведения сети, и в каждом узле КС могут храниться несколько параметров, отражающих его состояние. Однако, несмотря на то что SNMP является простым протоколом, у него имеются некоторые недостатки по оптимальному использованию пропускной способности сети. Например, опрос узлов сети в виде "один запрос, один ответ", необходимость идентификации объекта (OID — *object identifier*), невозможность фильтрации и т. д. увеличивают время отклика на запросы монитора, что особенно актуально в низкоскоростных сетях.

2. Системы поллинга

Системы поллинга являются разновидностью систем массового обслуживания, состоящих из нескольких очередей, которые обслуживает один сервер. При этом сервер по определенному правилу переключается между очередями и обслуживает

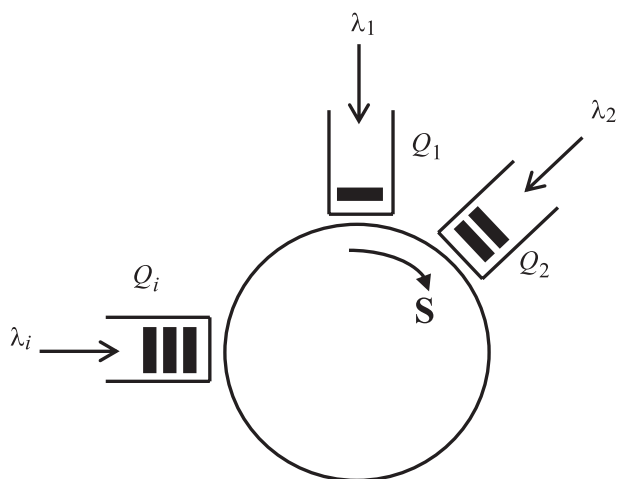


Рис. 1. Модель циклического поллинга

находящиеся в них заявки (рис. 1) [1]. Имеется большое количество литературы по исследованию систем поллинга, и в работах [21—24] приводится анализ систем поллинга с различными дисциплинами обслуживания, например с исчерпывающей (*exhaustive*), шлюзовой (*gated*) и ограниченной (*limited*) дисциплинами.

Термин "поллинг" происходит от так называемой схемы опроса, предназначенной для управления общим каналом передачи данных. В этой схеме опроса центральный компьютер (сервер) опрашивает каждый терминал (очередь) на наличие у него (клиента) какой-либо информации для передачи. При этом получивший запрос терминал передает информацию, а затем сервер переключается на следующий терминал, чтобы уже у него проверить наличие информации для передачи. В итоге сервер опрашивает все терминалы и таким образом управляет процессом передачи информации по общему каналу передачи данных.

Приведенная выше модель может быть рассмотрена как основная стандартная модель поллинга. Предполагается, что модель поллинга состоит из N очередей ($N \geq 1$) Q_1, \dots, Q_N и одного сервера S , циклически опрашивающего очереди по определенной дисциплине $D_i, i = 1, \dots, n$. При этом для переключения сервера S от очереди Q_i к очереди Q_{i+1} затрачивается определенное время, которое называется временем переключения и является независимой случайной величиной. В работе [25] приводится анализ поллинг-моделей без времени переключения, в которой предполагается, что переключения сервера происходят мгновенно.

Модель поллинга аналогична модели синхронного мультиплексирования с временным разделением (СМВР) [26]. Однако модель СМВР отличается тем, что, независимо от того, имеются ли в очередях заявки или нет, для обслуживания каждой очереди сервером выделяется фиксированное время.

Системы поллинга характеризуются процессом поступления в очереди заявок, процессом обслуживания заявок, процессом переключения сервера между очередями, размером буфера и дисциплиной обслуживания.

Процесс поступления заявок происходит в соответствии с их очередностью и независимо друг от друга. Часто в литературе эти процессы считаются пуассоновскими процессами с интенсивностью λ_i , где $i = 1, \dots, N$, т. е. временной интервал между заявками одного и того же типа имеет экспоненциальное распределение. Однако в некоторых реальных системах это утверждение может быть несостоятельным, и поэтому, исходя из сущности сбора данных мониторинга, мы рассматриваем поступление заявок с обновлением.

Процесс обслуживания заключается в том, что каждый раз, когда сервер опрашивает очередь, начинается период обслуживания, в котором обслуживаются заявки текущей очереди. Для обслуживания каждой заявки требуется определенное время. При этом предполагается, что для обслуживания одного и того же типа заявок время обслуживания является независимой, равномерно распределенной случайной величиной и не зависит от состояния системы.

Процесс переключения сервера между очередями заключается в переключении сервера с одной очереди на другую, и обычно для этого требуется некоторое время. Затрачиваемое на переключение между выбранной парой очередей время является независимой случайной распределенной величиной.

Несмотря на то что указанные стохастические временные показатели являются очень важными, они не позволяют полностью идентифицировать систему поллинга. Производительность систем поллинга также зависит от числа очередей, размера буфера, дисциплины обслуживания и дисциплины опроса.

Обычно размер буфера каждой очереди в системе поллинга предполагается неограниченным (бесконечным), в некоторых исследованиях рассматриваются также единичные буферы. Несмотря на то что предположения бесконечности не всегда совпадают с действительностью, считается, что размер буфера не является ограничивающим фактором.

Дисциплины обслуживания. Дисциплина обслуживания определяется числом заявок, которые обслуживаются в течение одного посещения сервером очереди. Внутри очереди заявки обслуживаются в порядке, определяемом дисциплиной обслуживания, причем чаще всего в порядке поступления в очередь. В литературе предлагается множество различных дисциплин обслуживания очередей. Дисциплины обслуживания могут быть классифицированы как дисциплины обслуживания с ограниченным числом заявок, в которых число заявок, обслуживаемых сервером при текущем посещении очереди,

ограничивается, и как дисциплины с ограниченным временем, в которых на посещение сервером очереди налагаются временные ограничения.

Кроме того, дисциплины обслуживания можно разделить на так называемую исчерпывающую, при которой сервер обслуживает заявки до тех пор, пока очередь не опустеет, и шлюзовую [27—29], при которой сервер обслуживает лишь те заявки, которые находились в очереди на момент подключения к ней сервера. В исчерпывающей дисциплине обслуживания вновь прибывающие в очередь заявки могут быть потенциальными кандидатами на обслуживание сервером, а в шлюзовой дисциплине они таковыми не являются. Если сервер обслуживает только те заявки, которые находились в очереди в момент начала цикла, то говорят о глобально-шлюзовой дисциплине, анализ которой проведен в работе [30].

Путем объединения различных дисциплин обслуживания может быть получено множество гибридных дисциплин. Некоторые из этих дисциплин обслуживания изучались в литературе, например, вероятностно-ограниченная дисциплина обслуживания [31]; l_i -ограниченная дисциплина, где число заявок, которое может обслуживать сервер, ограничено числом l_i , $l_i \geq 1$; l_i -уменьшающая дисциплина, при которой сервер обслуживает заявки в очереди до тех пор, пока ее длина не станет на l меньше, чем была в момент подключения сервера, $l \geq 1$.

При вероятностной дисциплине обслуживания число заявок, которое может обслужить сервер в очереди Q_i , определяется значением дискретной случайной величины ξ_i , имеющей закон распределения d_i^j , $j \geq 1$, причем закон распределения может меняться при каждом посещении очереди. Примерами вероятностных дисциплин являются дисциплина Бернулли и биномиальная дисциплина.

При дисциплине Бернулли [32] первая заявка в очереди Q_i обслуживается с вероятностью 1, а каждая последующая — с заданной вероятностью p_i . С вероятностью $1 - p_i$ сервер покидает очередь. Для данной дисциплины закон распределения имеет вид $d_i^j = p^{j-1}$, $j \geq 1$. Биномиальная дисциплина [33], при которой случайная величина ξ_i (число заявок, которое может обслужить сервер в очереди Q_i) имеет биномиальное распределение с параметрами X_i и p_i , где X_i — число заявок в очереди Q_i в момент подключения к ней сервера; p_i — некоторое число, $0 < p_i \leq 1$. Для данной дисциплины закон распределения имеет вид $d_i^j = C_{X_i}^j p_i^j (1 - p_i)^{X_i - j}$, $j = \overline{1, X_i}$, $d_i^j = 0$ для $j > X_i$. Подробная классификация дисциплин обслуживания систем поллинга дана в работе [34], где также приведены соотношения между

числом заявок в системе при различных дисциплинах обслуживания.

Способы переключения. Кроме множества дисциплин обслуживания очередей, которые описывают поведение сервера при обслуживании очереди, также имеется несколько способов переключения (выбора очереди) сервера между очередями. Основной моделью поллинга является циклическая модель, при которой сервер посещает очереди в порядке $Q_1, Q_2, \dots, Q_N, Q_1, Q_2, \dots, Q_N, \dots$. При этом время, затрачиваемое сервером на обход очередей от Q_1 до Q_N , называется циклом. Вместе с тем основная циклическая модель имеет некоторые обобщения, такие как вероятностный, периодический и приоритетный поллинги.

Вероятностный поллинг заключается в том, что сервер с вероятностью p_i , $i = \overline{1, N}$, выбирает очередь Q_i для обслуживания, при этом $\sum_{i=1}^N p_i = 1$. Воз-

можен также и другой вариант выбора очереди с вероятностью p_{ij} , $i, j = \overline{1, N}$, после посещения очереди Q_i сервер переключается к Q_j , $\sum_{j=1}^N p_{ij} = 1$, $i = \overline{1, N}$.

Периодический поллинг заключается в том, что сервер обслуживает очереди в фиксированном

порядке, установленном по таблице поллинга ($T(1), T(2), \dots, T(M)$) длины M ($M \geq N$), $T(i) \in \{1, \dots, N\}$, $i = \overline{1, M}$, в которой каждая очередь обслуживается по крайней мере один раз. При этом предполагается, что таблица поллинга содержит номера всех очередей системы, и время, затрачиваемое сервером на обход очередей от $Q_{T(1)}$ до $Q_{T(M)}$, называется циклом. В работе некоторых систем поллинга выделяют гамильтоновы цикл — это время, за которое сервер посещает все очереди, причем ровно один раз. Частными случаями периодического порядка обхода очередей являются обход типа "звезда", когда очереди обслуживаются в порядке $Q_1, Q_2, Q_1, Q_3, \dots, Q_1, Q_N$, и элеваторный порядок (порядок лифта) обхода очередей, при котором очередь обслуживается в порядке $Q_1, Q_2, \dots, Q_{N-1}, Q_N, Q_{N-1}, \dots, Q_2, Q_1$.

Приоритетный поллинг заключается в том, что система имеет очереди разных приоритетов, и некоторая очередь может быть обслужена сервером, только если более приоритетные очереди не содержат заявки.

Если все очереди системы поллинга имеют дисциплины обслуживания одного типа, то говорят о системе поллинга с данной дисциплиной обслуживания (например, с исчерпывающей, ограниченной и другими дисциплинами обслуживания). Если дисциплины обслуживания очередей различные, то говорят о системе поллинга со смешанной

дисциплиной обслуживания. В зависимости от цели и задачи управления и мониторинга КС может быть выбрана та или иная дисциплина обслуживания.

3. Модель поллинга узлов КС

Поллинг широко используется в КС для определения состояния ее узлов в режиме реального времени. Обычно во многих сетях опрос проводится периодически с фиксированной частотой и для этого используется протокол SNMP.

Исходя из того, что мониторинг КС является процессом сбора данных характеристик узлов, его можно смоделировать как систему поллинга. Для мониторинга КС предлагается простейшая абстрактная модель (рис. 2), которая состоит из центрального узла, так называемого центра мониторинга сети или монитора M и N сетевых узлов, к которым относятся хосты, серверы, маршрутизаторы, коммутаторы, хабы, межсетевые экраны и т. д. Предложенная модель поллинга узлов КС основывается на протоколе SNMP. В этой модели монитор использует SNMP-поллинг и МІВ-базу для получения

данных мониторинга от сетевых узлов. Монитор получает информацию от МІВ-базы путем периодического опроса встроенных в сетевые узлы SNMP-агентов в режиме реального времени. Как правило, опрос проводится синхронно отправлением запросов к SNMP-агентам, находящимся в узлах, и получением информации о текущем состоянии узлов КС. Вместе с тем для точного определения состояния узлов в режиме реального времени опрос узлов должен проводиться часто, и при таком мониторинге небольших КС существенные проблемы не возникают. В то же время постоянное изменение содержания МІВ-таблиц приводит к необходимости частого опроса узлов, чтобы обеспечить актуальность собранных данных мониторинга. Однако с ростом размеров и сложности КС увеличивается и размер МІВ-таблиц (например, таблицы ТСР-соединений и IP-маршрутизации), что приводит к увеличению времени опроса. Но редкий опрос узлов затрудняет понимание их состояния в режиме реального времени, а также в целом состояние КС.

Из описанной выше модели мониторинга КС (см. рис. 2) можно заключить, что процесс опроса узлов сети монитором, т. е. процессы от отправки монитором запроса к сетевым узлам до получения и обработки данных мониторинга, имеют сходство с моделью системы поллинга (см. рис. 1).

Используя это сходство, модель опроса монитором сетевых узлов может быть показана в виде системы поллинга (рис. 3). При этом для оптимизации мониторинга КС следует учитывать порядок опроса узлов сети, число обслуживаемых заявок и порядок обслуживания заявок в каждой очереди.

Математическая модель мониторинга КС (модель поллинга узлов КС) может быть описана следующим образом. Система имеет один сервер (монитор) M и N очередей (узлов сети) Q_1, \dots, Q_N с неограниченным числом мест (с неограниченным буфером) для ожидания (неограниченное число заявок может быть сохранено в каждой очереди без потерь). Предполагается, что единственный монитор M обслуживает по циклу N узлов сети, переключаясь с одного узла на другой. Рассматривается система поллинга с исчерпывающей дисциплиной обслуживания очередей, т. е. при текущем посещении сервером очереди Q_i обслуживаются все заявки (обрабатываются все данные мониторинга), находящиеся в ней, и с вероятностным переключением сервера с очереди на очередь с вероятностью $p_i, i = \overline{1, N}$, что позволит снизить объем трафика поллинга. При этом полагается, что внутри очереди заявки обслуживаются в порядке поступления, и процесс поступления заявок (данных мониторинга) является независимым пуассоновским процессом. Заявки поступают в очередь Q_i с интенсивностью λ_i , где $i = 1, \dots, N$, и общая интенсивность поступления

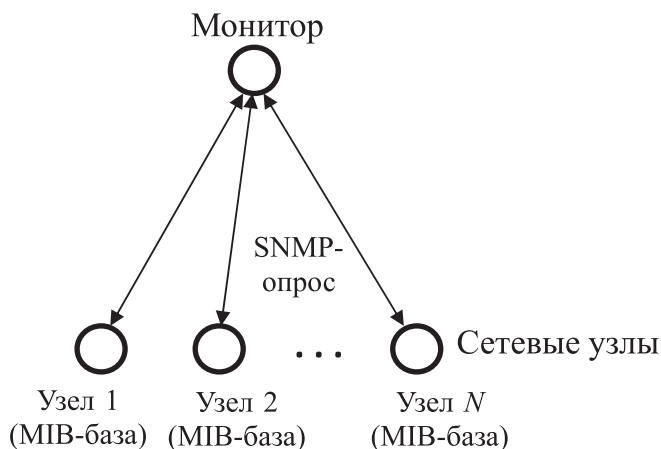


Рис. 2. Абстрактная модель мониторинга КС

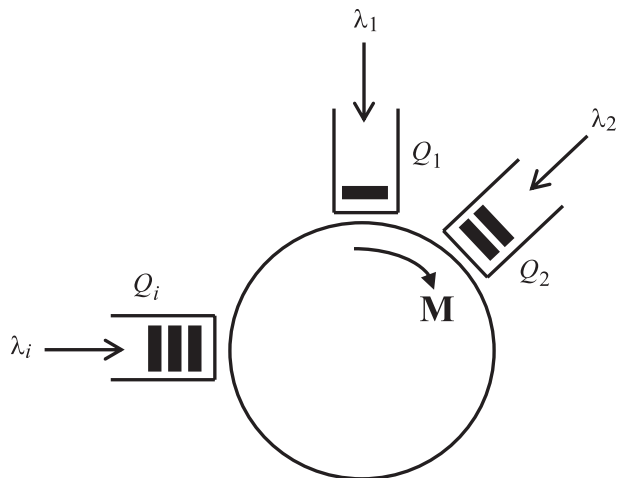


Рис. 3. Модель поллинга монитором сетевых узлов

заявок в систему равна $\lambda = \sum_{i=1}^N \lambda_i$. Заявки, поступающие к очереди Q_i , называются заявками i -класса,

и время их обслуживания является независимой случайной величиной со средним значением β_i и

вторым моментом $\beta_i^{(2)}$, $i = 1, \dots, N$. Загрузка очереди Q_i обозначается через ρ_i и определяется как $\rho_i := \lambda_i \beta_i$, где $i = 1, \dots, N$ и общая загрузка системы

$\rho := \sum_{i=1}^N \rho_i$. Время обслуживания очереди Q_i является независимой и равномерно распределенной

величиной с функцией распределения $B_i(t)$ с моментами $b_i^{(r)} = \int_0^{\infty} t^r dB_i(t)$, $r \geq 1$, и преобразованием

Лапласа—Стилтьеса (*Laplace—Stieltjes transform (LST)*) $\beta_i(s) = \int_0^{\infty} e^{-st} dB_i(t)$, $i = \overline{1, N}$. При этом полагается,

что потоки заявок и время обслуживания заявок являются независимыми. По классификации Кендалла такая система называется системой поллинга с очередями типа $M/GI/1$. Необходимым и достаточным условием стабильности (эргодичности) описываемой системы считается $\rho < 1$ [35], и это условие не зависит от времени переключения сервера из очереди в очередь, причем система является незагруженной.

Время ожидания заявки в очереди Q_i определяется временем между поступлением произвольной заявки в систему и началом ее обслуживания. Предполагается, что в очереди заявки обрабатываются по принципу "первым пришел — первым обслужен" (*First Come, First Served (FCFS)*).

Сервер после обслуживания покидает очередь Q_i , используя время переключения типа i , которое является независимой случайной величиной со средним значением t_i , вторым моментом $t_i^{(2)}$, $i = 1, \dots, N$, и преобразованием Лапласа—Стилтьеса $T_i(s)$, $i = \overline{1, N}$. При этом предполагается, что время переключения сервера не зависит от процессов поступления и обслуживания заявок в очереди Q_i .

Среднее время цикла, осуществляемого сервером, как правило, не зависит от очереди Q_i и обозначается как $E[C]$. Среднее время ожидания обслуживания заявки в очереди Q_i обозначается как $E[W_i]$. В работе [36] дифференцированием преобразования Лапласа—Стилтьеса распределения времени ожидания заявок для $E[W_i]$ найдены следующие выражения:

для исчерпывающей дисциплины

$$E[W_i] = (1 - \rho_i) \frac{E[C_i^{*2}]}{2E[C]};$$

для шлюзовой дисциплины

$$E[W_i] = (1 - \rho_i) \frac{E[C_i^{*2}]}{2E[C]}.$$

4. Оптимизация поллинга узлов КС

Исходя из описанной в предыдущем разделе модели поллинга узлов КС, задачу оптимизации процесса мониторинга КС сводим к задаче оптимизации модели поллинга узлов КС. Как было показано выше, системы поллинга в основном характеризуются процессом поступления в очереди заявок, процессом обслуживания заявок, процессом переключения сервера между очередями. Причем каждый процесс характеризуется конкретным параметром, например, процесс поступления характеризуется интенсивностью поступления заявок (данных мониторинга), процесс обслуживания заявок — средним временем ожидания заявок в очереди, а процесс переключения сервера между очередями — временем переключения сервера между очередями. При этом для оптимизации процесса мониторинга КС необходимо динамическое управление поступлением потока данных мониторинга в монитор и дисциплиной обслуживания на основании параметров состояния входного потока, таких как статический приоритет, критичное время жизни заявки, размер штрафа за потерю заявки и т. д. При этом основной задачей является обеспечение в режиме реального времени непрерывности поступления данных мониторинга в монитор и их актуальности. Однако в данной работе вопросы штрафа за потерю заявок не рассматриваются.

Обеспечение непрерывности и актуальности данных мониторинга в режиме реального времени возможно при минимизации среднего времени пребывания заявок в очередях. При этом среднее время пребывания заявки в очереди Q_i может быть определено следующим выражением:

$$E[TS_i] = E[W_i] + \beta_i,$$

где $E[W_i]$ — среднее время ожидания обслуживания заявки в очереди Q_i ; β_i — среднее время обслуживания заявки в очереди Q_i .

Если порог времени пребывания заявки в очереди Q_i (или критичное время жизни заявки) обозначим через TS_i^* , то для минимизации времени пребывания заявки в очереди Q_i необходимо, чтобы выполнялось условие $E[TS_i] \leq TS_i^*$. Учитывая то, что время обслуживания является независимой случайной величиной, минимизация среднего времени пребывания заявки в очереди Q_i может быть достигнута минимизацией среднего времени ожидания обслуживания заявки $E[W_i]$. При этом с минимизацией времени ожидания обслуживания приоритет заявки может быть повышен, поскольку со временем вероятность превышения заявкой порога TS_i^* может увеличиваться. Однако не всегда удается по-

лучить явные формулы для вычисления среднего времени ожидания заявок в каждой очереди и в работах по исследованию систем поллинга обычно находят приближенные формулы. Иногда задача нахождения средних времен ожидания сводится к нахождению взвешенной суммы этих показателей.

Для оптимизации рассмотренной выше модели поллинга узлов КС используем полученные в работе [37] оптимальные значения вероятностей p_i^* , $i = \overline{1, N}$, для системы поллинга с исчерпывающей или шлюзовой дисциплиной обслуживания очередей и с вероятностной дисциплиной задаваемой вероятности p_i , $i = \overline{1, N}$, минимизирующей взвешенную сумму средних времен ожидания заявок в очереди Q_i :

$$\sum_{i=1}^N \rho_i E[W_i] = \rho \frac{\sum_{i=1}^N \lambda_i \beta_i^{(2)}}{2(1-\rho)} - \frac{t_i}{1-\rho} \sum_{k \in e} \frac{\rho_k^2}{p_k} + \frac{t_i}{1-\rho} \sum_{k=1}^N \frac{\rho_k}{p_k} - \rho t_i + \rho \frac{t_i^{(2)}}{t_i},$$

где e — множество номеров очередей с исчерпывающим обслуживанием.

Задача оптимизации заключается в следующем:

$$\sum_{i=1}^N \rho_i E[W_i] \xrightarrow{p_1, \dots, p_N} \min;$$

$$\sum_{i=1}^N p_i = 1, p_1 \geq 0, \dots, p_N \geq 0.$$

По сути, она является задачей оптимальной маршрутизации сервера при переключении между очередями, и нетрудно видеть, что минимизация этой целевой функции эквивалентна минимизации загрузки системы.

Несмотря на то что неявное выражение для взвешенной суммы средних времен ожидания заявок в очереди Q_i известно, его непосредственная оптимизация невозможна. Эта задача является классической задачей нелинейной оптимизации с линейными ограничениями, которая решается с помощью метода множителей Лагранжа. Решение этой задачи имеет следующий вид [38]:

$$p_k^* = \frac{\sqrt{\rho_k(1-\rho_k)}}{\sum_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum_{j \in g} \sqrt{\rho_j}}, k \in e;$$

$$p_k^* = \frac{\sqrt{\rho_k}}{\sum_{j \in e} \sqrt{\rho_j(1-\rho_j)} + \sum_{j \in g} \sqrt{\rho_j}}, k \in g,$$

где g является множеством номеров очередей с шлюзовым обслуживанием.

Современные КС являются настолько сложными и неопределенными системами, что это затрудняет их описание в виде детерминированных моделей.

В работе предлагается методология повышения эффективности мониторинга КС, которая позволит уменьшить время, затрачиваемое на мониторинг, и обеспечит непрерывность обновления и актуальность данных мониторинга. Для этого используется модель системы поллинга и предлагается модель поллинга узлов КС.

Предполагается, что основным показателем оптимальности систем поллинга является минимизация среднего времени ожидания заявок в очереди. Поэтому задача оптимизации процесса мониторинга КС была сведена к задаче оптимизации модели поллинга узлов КС посредством минимизации среднего времени ожидания заявок в очереди.

Список литературы

1. **Levy H., Sidi M.** Polling systems: applications, modeling, and optimization. IEEE Transactions on Communications, 1990, vol. 38, no. 10, pp. 1750–1760.
2. **Network Monitoring and Analysis. A Protocol Approach to Troubleshooting** / Ed. Wilson. New Jersey: Prentice Hall, Inc. Upper Saddle River, 2002. 350 p.
3. **Gao F., Gutierrez J.** A Dynamic Polling Scheme for the Network Monitoring Problem // Proc. of the 7th Asia-Pacific Network Operations and Management Symposium (APNOMS 2003), Fukuoka, Japan, October, 2003. P. 489–500.
4. **Fumio Machida, Masahiro Kawato, Yoshiharu Maeno.** Polling Schedule Optimization for Adaptive Monitoring to Scalable Enterprise Systems // International Journal On Advances in Intelligent Systems. 2008. Vol. 1, no. 1. P. 11–22.
5. **Park S., Park M.** An efficient transmission for large MIB tables in polling-based SNMP // Proc. of the 10th International Conference on Telecommunications. 2003. Vol. 1. P. 246–252.
6. **Скатков А. В., Воронин Д. Ю., Данильчук Д. Н.** Система поддержки принятия решений по выбору структуры системы мониторинга // Технология и конструирование в электронной аппаратуре. 2008. № 2. С. 10–13.
7. **Шыхалиев Р. Г.** Об одном методе сокращения размерности анализируемых признаков сетевых трафиков, используемых для мониторинга компьютерных сетей // Телекоммуникации. 2011. № 06. С. 44–48.
8. **Breitbart Y., Dragan F., Gobjuka H.** Effective Network Monitoring // Proc. of the International Conference On Computer Communications and Networks (ICCCN 2004). October 11–13, 2004. Chicago, IL, USA. P. 394–399.
9. **Breitbart Y., Chan C. Y., Garofalakis M., Rastogi R., Silberschatz A.** Efficiently Monitoring Bandwidth and Latency in IP Networks // Proc. of IEEE INFOCOM, 2000.
10. **Breitbart Y., Dragan F., Gobjuka H.** Effective Monitor Placement in Internet Networks // Journal of Networks. 2009. Vol. 4, no. 7. P. 657–666.
11. **Leppinen M., Pulkkinen P., Rautiainen A.** Java- and CORBA-based network management // Computer. 1997. Vol. 30, no. 6. P. 83–87.
12. **Lee J. O.** Enabling network management using Java technologies // IEEE Communications Magazine. 2000. Vol. 38, no. 1. P. 116–123.
13. **Gavalas D., Greenwood D., Ghanbari M., O'Mahony M.** Advanced network monitoring applications based on mobile/intelligent agent // Computer Communications. 2000. Vol. 23, no. 8. P. 720–730.
14. **Romascanu D., Zilbershtein I. E.** Switch monitoring — The new generation of monitoring for local area networks // Bell Labs Technical Journal. 1999. Vol. 4, no. 4. P. 42–54.

15. **Shaffi A. S., Al-Obaidy M.** Managing network components using SNMP // International Journal of Scientific Knowledge. 2013. Vol. 2, no. 3. P. 11–18.

16. **Thottan M. K.** SEQUIN: An SNMP-based MPLS network monitoring system // Bell Labs Technical Journal. 2003. Vol. 8, no. 1. P. 95–111.

17. **Mangaiyarkarasi R., Sivakumar S. R.** Nnetwork monitoring using snmp protocol // International Journal of Power Control Signal and Computation (IJPCSC). Jan-Mar 2012. Vol. 3, no. 1.

18. **Case J., Fedor M., Schoffstall M., Davin J.** Simple Network Management Protocol (SNMP), RFC 1157. Network Working Group, IETF, 1990.

19. **Stallings W.** SNMP, SNMPv2, SNMPv3, and RMON 1 and 2. Addison Wesley, 1996.

20. **McCloghrie K., Rose M.** Management information base for network management of TCP/IP-based internets: MIB-II, RFC 1213, March 1991.

21. **Вишнеvский В. М., Семенова О. В.** Математические методы исследования систем поллинга // Автоматика и телемеханика. 2006. № 2. С. 3–56.

22. **Вишнеvский В. М., Дудин А. Н., Клименок В. И., Семенова О. В., Шпилев С. А.** Система со шлюзовым обслуживанием и отдыхами прибора, зависящими от состояния системы // Четвертая Международная конференция по проблемам управления (МКПУ-IV) 26–30 января 2009: Сб. трудов. М.: Изд. ИПУ РАН, 2009. С. 1767–1772.

23. **Семенова О. В., Петроченков В. В.** Система поллинга с нетерпеливыми заявками и исчерпывающей дисциплиной обслуживания // Distributed Computer and Communication Networks. Theory and Applications (DCCN-2008), М.: ИТПРАС, 2008. С. 45–56.

24. **Takagi H.** Queueing analysis of polling models: an update. // Stochastic Analysis of Computer and Communication Systems / Ed. H. Takagi. Amsterdam: North-Holland Publ. Co., 1990. P. 267–318.

25. **Borst S. C., Boxma O. J.** Polling systems with and without switch-over times // Operations Research. 1997. Vol. 45, no. 4. P. 536–543.

26. **Forouzan I., Behrouz A.** Data communications and networking. — 4th ed. 2007, 1134 p.

27. **Van der Mei R. D., Resing J. A. C.** Analysis of polling models with two-stage gated service: fairness versus efficiency // Mason L., Drwiega T., Yan J. (Eds.). Managing Traffic Performance in Converged Networks — the Interplay between Convergent and Divergent Forces, ITC2007, Lecture Notes in Computer Science 4516, 2007. P. 544–555.

28. **Van der Mei R. D., Resing J. A. C.** Polling systems with two-phase gated service: heavy traffic results for the waiting-time distributions // Probability in the Engineering and Informational Sciences. 2008. Vol. 4, no. 22. P. 623–651.

29. **Van der Mei R. D., Roubos A.** Polling models with multi-phase gated service // Operations Research. 2012. Vol. 1, no. 198. P. 25–56.

30. **Leung K. K.** Cyclic-service systems with probabilistically-limited service // IEEE Journal on Selected Areas in Communications. 1991. Vol. 2, no. 9. P. 185–193.

31. **Boxma O. J., Levy H., Yechiali U.** Cyclic reservation schemes for efficient operation of multiple-queue single-server systems // Annals of Operations Research. 1992. Vol. 3, no. 35. P. 187–208.

32. **Keilson J., Servi L. D.** Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules // Journal of Applied Probability. 1986, no. 23. P. 790–802.

33. **Levy H.** Analysis of cyclic polling systems with binomial gated service. In: Hasegawa T., Takagi H., Takahashi Y. (Eds.), Performance of Distributed and Parallel Systems. P. 127–139. Amsterdam: North-Holland, 1989.

34. **Levy H., Sidi M., Boxma O. J.** Dominance relations in polling systems // Queueing Systems. 1990. Vol. 6, is. 1. P. 155–171.

35. **Takagi H.** Analysis of polling systems. (Cambridge, Mass. MIT Press), 1986. 197 p.

36. **Boon M. A. A., Van der Mei R. D., Winands E. M. M.** Applications of polling systems // Surveys in Operations Research and Management Science. 2011. Vol. 16, is. 2. P. 67–82.

37. **Boxma O. J.** Static optimization of queueing systems // Recent Trends in Optimization Theory and Applications / Ed. Agwal R. P. Singapore: World Scientific Publ., 1995. P. 1–16.

38. **Boxma O. J., Levy H., Weststrate J.A.** Optimization of polling systems // Performance '90, eds. P. J. B. King, I. Mitrani, R. J. Pooley. North-Holland, Amsterdam. P. 349–361.

R. H. Shikhaliyev, PhD. on technical sciences, Leading Researcher

Institute for Information Technologies, Azerbaijan National Academy of Sciences, Baku, Azerbaijan
ramiz@science.ab.az

Improving the Monitoring Efficiency of Computer Networks Based on the Polling System Optimization

It is proposed the optimization model for online monitoring of computer networks (CN), which allow minimize of monitoring time. The main goal is to optimize the monitoring of CN nodes at given network resources. For achievement of this purpose it is offered to use model of polling system, by optimizing which it is possible to reach optimization of CN monitoring. The literature contains a number of works, which are devoted to improve the efficiency of network monitoring through the optimization of certain aspects of the polling. The analysis shows, that despite the large contribution to improving the efficiency of network monitoring, these works are based on the architectural solution of the problem of optimization of certain aspects of the polling. And in this works not considered characteristic of the polling system such as the intensity of arriving queries (monitoring data), the mean waiting time in the queue of requests, the server switching time between queues and etc. For this reason, in this paper to improve the efficiency of CN monitoring is proposed to optimize the polling system itself. As one of the main indicators of polling systems optimality is minimal mean waiting time of queries in each queue. Therefore, for optimization of the CN nodes polling, we proposed to minimize of mean waiting time for queries in queues. In our opinion it would be a unified approach to increase the efficiency of CN monitoring based on the polling system.

Keywords: computer networks, network monitoring, network nodes polling, SNMP-polling, polling systems, polling discipline, mean waiting time

References

1. **Levy H., Sidi M.** Polling systems: applications, modeling, and optimization. *IEEE Transactions on Communications*, 1990, vol. 38, no. 10, pp. 1750–1760.
2. **Network Monitoring and Analysis. A Protocol Approach to Troubleshooting.** Ed. Wilson. New Jersey, Prentice Hall, Inc. Upper Saddle River, 2002, 350 p.
3. **Gao F., Gutierrez J.** A Dynamic Polling Scheme for the Network Monitoring Problem. *Proc. of the 7th Asia-Pacific Network Operations and Management Symposium (APNOMS 2003)*, Fukuoka, Japan, October, 2003, pp. 489–500.
4. **Machida F., Kawato M., Maeno Y.** Polling Schedule Optimization for Adaptive Monitoring to Scalable Enterprise Systems. *International Journal On Advances in Intelligent Systems*, 2008, vol. 1, no. 1, pp. 11–22.
5. **Park S., Park M.** An efficient transmission for large MIB tables in polling-based SNMP. *10th International Conference on Telecommunications*, 2003, vol. 1, pp. 246–252.
6. **Skatkov A. V., Voronin D. Yu., Danil'chuk D. N.** Sistema podderzhki prinyatiya reshenii po vyboru struktury sistemy monitoringa. *Tekhnologiya i konstruirovaniye v elektronnoy apparature*, 2008, no. 2, pp. 10–13.
7. **Shikhaliyev R. H.** Ob odnom metode sokrashcheniya razmernosti analiziruemykh priznakov setevykh trafikov, ispol'zuemykh dlya monitoringa komp'yuternykh setei. *Telekommunikatsii*, 2011, no. 6, pp. 44–48.
8. **Breitbart Y., Dragan F., Gobjuka H.** Effective Network Monitoring. *Proc. of the International Conference on Computer Communications and Networks (ICCCN 2004)*, 2004, October 11–13, Chicago, IL, USA, pp. 394–399.
9. **Breitbart Y., Chan C. Y., Garofalakis M., Rastogi R., Silberschatz A.** Efficiently Monitoring Bandwidth and Latency in IP Networks. *Proc. of the Conference on Computer Communications (IEEE INFOCOM 2001)*, Anchorage, Alaska, USA, April 22–26, 2001, vol. 2, pp. 933–942.
10. **Breitbart Y., Dragan F., Gobjuka H.** Effective Monitor Placement in Internet Networks. *Journal of Networks*, 2009, vol. 4, no. 7, pp. 657–666.
11. **Leppinen M., Pulkkinen P., Rautiainen A.** Java- and CORBA-based network management. *Computer*, 1997, vol. 30, no. 6, pp. 83–87.
12. **Lee J. O.** Enabling network management using Java technologies. *IEEE Communications Magazine*, 2000, vol. 38 no. 1, pp. 116–123.
13. **Gavalas D., Greenwood D., Ghanbari M., O'Mahony M.** Advanced network monitoring applications based on mobile/intelligent agent. *Computer Communications*, 2000, vol. 23 no. 8, pp. 720–730.
14. **Romascanu D., Zilbershtein I. E.** Switch monitoring — The new generation of monitoring for local area networks. *Bell Labs Technical Journal*, 1999, vol. 4, no. 4, pp. 42–54.
15. **Shaffi A. S., Al-Obaidy M.** Managing network components using SNMP. *International Journal of Scientific Knowledge*, 2013, vol. 2, no. 3, pp. 11–18.
16. **Thottan M. K.** SEQUIN: An SNMP-based MPLS network monitoring system. *Bell Labs Technical Journal*, 2003, vol. 8 no. 1, pp. 95–111.
17. **Mangaiyarkarasi R., Siyakumar S. R.** Nnetwork monitoring using snmp protocol. *International Journal of Power Control Signal and Computation (UPCSC)*, Jan–Mar 2012, vol. 3, no. 1.
18. **Case J., Fedor M., Schoffstall M., Davin J.** *Simple Network Management Protocol (SNMP)*, RFC 1157, Network Working Group, IETF, 1990.
19. **Stadligns W.** SNMP, SNMPv2, SNMPv3, and RMON 1 and 2, 1996, Addison Wesley.
20. **McCloghrie K., Rose M.** Management information base for network management of TCP/IP-based internets: MIB-II, RFC 1213, March 1991.
21. **Vishnevskii V. M., Semenova Q. V.** Matematicheskie metody issledovaniya sistem pollinga. *Avtomatika i telemekhanika*, 2006, no. 2, pp. 3–56.
22. **Vishnevskii V. M., Dudin A. N., Klimenok V. I., Semenova O. V., Shipilev S. A.** Sistema so shlyuzovym obsluzhivaniem i otdykhami pribora, zavislyashchimi ot sostoyaniya sistemy. *Chetvertaya Mezhdunarodnaya konferentsiya po problemam upravleniya (MKPU-IV): Sbornik trudov.* M.: IPU RAN, January 26–30, 2009, pp. 1767–1772.
23. **Semenova O. V., Petrochenkov V. V.** Sistema pollinga s neterpelivymi zayavkami i ischerpyvayushchei distsiplinoi obsluzhivaniya. *Distributed Computer and Communication Networks. Theory and Applications (DCCN-2008)*, Moscow: IITP RAS, 2008, pp. 45–56.
24. **Takagi H.** Queueing analysis of polling models: an update. In: H. Takagi (ed.), *Stochastic Analysis of Computer and Communication Systems*. North-Holland Publ. Cy., Amsterdam, 1990, pp. 267–318.
25. **Borst S. C., Boxma O. J.** Polling systems with and without switch-over times. *Operations Research*, 1997, vol. 45, no. 4, pp. 536–543.
26. **Forouzan I., Behrouz A.** *Data communications and networking.* — 4th ed., Published by McGraw-Hill, 2007, 1134 p.
27. **Van der Mei R. D., Resing J. A. C.** Analysis of polling models with two-stage gated service: fairness versus efficiency. In: L. Mason, T. Drwiega, and J. Yan (Eds.), *Managing Traffic Performance in Converged Networks — the Interplay between Convergent and Divergent Forces, ITC2007, Lecture Notes in Computer Science 4516*, 2007, Berlin, Springer, pp. 544–555.
28. **Van der Mei R. D., Resing J. A. C.** Polling systems with two-phase gated service: heavy traffic results for the waiting-time distributions. *Probability in the Engineering and Informational Sciences*, 2008, vol. 4, no. 22, pp. 623–651.
29. **Van der Mei R. D., Roubos A.** Polling models with multi-phase gated service. *Operations Research*, 2012, vol. 1, no. 198, pp. 25–56.
30. **Leung K. K.** Cyclic-service systems with probabilistically-limited service. *IEEE Journal on Selected Areas in Communications*, 1991, vol. 2, no. 9, pp. 185–193.
31. **Boxma O. J., Levy H., Yechiali U.** Cyclic reservation schemes for efficient operation of multiple-queue single-server systems. *Annals of Operations Research*, 1992, vol. 3, no. 35, pp. 187–208.
32. **Keilson J., Servi L. D.** Oscillating random walk models for GI/G/I vacation systems with Bernoulli schedules. *Journal of Applied Probability*, 1986, no. 23, pp. 790–802.
33. **Levy H.** Analysis of cyclic polling systems with binomial gated service. In: Hasegawa T., Takagi H., Takahashi Y. (Eds.), *Performance of Distributed and Parallel Systems*. North-Holland, Amsterdam, 1989, pp. 127–139.
34. **Levy H., Sidi M., Boxma O. J.** Dominance relations in polling systems. *Queueing Systems*, 1990, vol. 6, is. 1, pp. 155–171.
35. **Takagi H.** *Analysis of polling systems.* Cambridge, Mass., MIT Press, 1986, 197 p.
36. **Boon M. A. A., Van der Mei R. D., Winands E. M. M.** Applications of polling systems. *Surveys in Operations Research and Management Science*, 2011, vol. 16, issue 2, pp. 67–82.
37. **Boxma O. J.** Static optimization of queueing systems. *Recent Trends in Optimization Theory and Applications.* Ed. Agwal R. P. Singapore: World Scientific Publ., 1995, pp. 1–16.
38. **Boxma O. J., Levy H., Weststrate J. A.** Optimization of polling systems. *Perfomence'90*, eds. P. J. B. King, I. Mitrani, R. J. Pooley, North-Holland, Amsterdam, pp. 349–361.