

基于 Jmeter 的智能法律问答系统性能测试

杨莹

(中国软件与技术服务股份有限公司, 北京 100081)

摘要: 随着智能法律问答系统近年的应用逐步扩大, 对此类系统进行充分有效的性能测试显得十分重要; 分析此类系统的设计与使用特点, 性能测试的主要难点包括: 新型协议的模拟, 数据证书的认证, 压力机分布式部署, 测试数据的选取, 测试结果的校验, 性能业务场景的设置; 面对以上这些难点, 选用 Jmeter 工具, 利用其对双向认证的 TLS1.2 协议证书配置和 WebSocket 协议, 压力机分布式部署的支持, 完成性能测试环境的搭建; 同时搜集常用的法律问题作为测试数据, 分析此类系统典型的应用场景, 并加入对问题答案的校验功能, 增加了性能测试的科学性和有效性。

关键词: 智能问答; 性能测试; 安全证书认证; 分布式部署; 数据匹配

Performance Testing of Intelligent Legal Question Answering System Based on Jmeter

Yang Ying

(China Software and Technical Service Co. Ltd, Beijing 100081, China)

Abstract: It is very important to the full performance testing for intelligent legal question answering system with the gradual popularization of such systems in recent years. It includes many difficulties for this performance testing, for example, simulation of new protocol, authentication of data certificate, distributed deployment of press machine, selection of test data, verification of test results, setting of performance business scenarios. This paper chooses Jmeter tool to complete the construction of performance testing environment, because it supports the two-way authentication TLS1.2 protocol, WebSocket protocol and distributed deployment of multiple press machine. At the same time, through collecting common legal questions as test data, analyzing typical application scenarios of such systems, and adding the function of checking the answers to the questions, these increase the scientificity and effectiveness of performance testing.

Keywords: intelligent question answering; security certificate authentication; distributed deployment; data match

0 引言

法律智能问答系统^[1]由于其专业知识强、涉及领域广、用户群体大、使用便捷、节约人力等优点, 近些年被各级法院积极用于满足人民群众对法律知识的咨询需求。此类系统设计^[2-3]上对支持的用户访问量, 存储的法律知识量, 检索速度, 数据安全都提出较高要求。采用 TLS1.2 双向认证并混合使用 http 协议与 WebSocket 协议恰好满足 Web 法律智能问答系统以上几点要求。在信息推送机制^[4]上, 作为较新的 WebSocket 协议相比于 Http 协议, 建立一次消息连接就可以实现多次消息传输的机制很大程度地提高了网络与服务器资源的利用率。采用 TLS1.2 协议对 WebSocket 协议进行加密所形成的 wss 协议 (即 TLS 协议加密的 WebSocket 协议), 又进一步保证了传输数据的保密性和完整性, 提高互联网数据传输的安全性。针对这种新协议设计的系统的性能测试, 在测试工具选取, 测试脚本编制, 性能测试场景设计几方面必然会与传统的基于 http 协议的 Web 应用系统的性能测试有明显的区别, 本文将围绕以上

几个方面系统地阐述此类 Web 法律智能问答系统^[5]进行性能测试的方法。

1 工具选择

目前国内主流性能测试工具是 Loadrunner 与 Jmeter^[6]。表 1 从不同角度分析了这两种工具对 Web 法律智能问答系统进行性能测试的支持情况。

从表 1 的对比分析来看, 两种工具同时满足此类系统性能测试的情况下, 从支持的并发用户数量, 安装便捷性及测试成本方面比较, Jmeter3.1 更适合作为此类系统的性能测试工具。因此, 下文将选用 Jmeter3.1 工具讲述性能测试环境的搭建。

2 测试工具配置

2.1 向 Jmeter 添加 WebSocket 协议测试所需插件

Jmeter 本身并不支持对 Html5 新增的 WebSocket 协议的测试, 但它可以通过安装插件的方式达到对 WebSocetk 性能测试的支持。为了支持 wss 协议测试, 需要安装的插件如下:

JMeterWebSocketSamplers-0.11.1.jar: 插件中包含的 WebSocket Open Connection、WebSocket Request-Response Sampler、WebSocket Close 三个组件。

收稿日期:2019-05-21; 修回日期:2019-06-22。

作者简介:杨莹(1983-),女,吉林四平人,工程硕士,工程师,主要从事软件工程、计算机应用和自动化测试方向的研究。

表 1 性能测试工具比较

工具名称	Jmeter	Loadrunner
是否支持 WebSocket	支持	支持
是否支持 TLS1.2	支持	支持
支持以上两种协议的 工具版本	3.1 及以上	12 及以上
支持并发用户数	无限制	正版的并发用户数 受购买的证书权限 限制, 试用版支持 并发用户 50
安装便捷性	程序大小 48M, 无需安 装直接解压使用, 但需 要电脑安装 java1.7 及 以上	安装包大小 1.24G, 安装过程自动安装 所需组件
插件扩展性	支持安装 java 插件	不支持插件扩展
分布式集群部署	支持	支持
支持操作系统	Windows, Linux, Mac	Windows, Unix

可以支持 wss 协议测试, 将此 jar 包放置到 Jmeter 目录的 \lib\ext\ 目录下, 重新启动 jmeter 可以在 sampler 中看到以上几个组件。

2.2 双向认证的客户端证书创建与配置

TLS 协议分为单向认证与双向认证。单向认证, TLS 握手阶段只验证服务器证书的真实性。双向认证, TLS 握手阶段客户端与服务器需要验证对方证书的真实性。基于对 Jmeter 的 TLS1.2 协议的支持, 从 Jmeter 的源代码分析 HTTPHC4Impl 类中创建 https 调用的是 LazySchemeSocketFactory 类, 此类又是调用 HC4TrustAllSSLSocketFactory 类, 这个类中有明确注释对 X509Certificate 的证书全部信任, Jmeter 对于服务器端证书采用全部信任, 因此对于单向认证的情况, 无需单独创建证书文件, 使用 Jmeter 自动创建的证书即可。而对于双向认证, Jmeter 需要配置客户端证书。从 PC 端浏览器导出客户端证书默认是 cer 格式, 移动端使用的是 p12 格式, 这两种格式的证书要被 Jmeter 使用, 都需要先转换为 jks 或 keystore 格式。

2.2.1 转换证书格式

转换方法如下:

使用 java 的 bin 目录下的 keytool 工具, 然后分别使用如下命令:

```
keytool -importcert -file 原证书路径\原证书文件名.cer -alias 新证书别名 -keystore 新证书路径\新证书文件名.jks
```

```
keytool -importkeystore -srckeystore 原证书路径\原证书文件名.p12 -srcstoretype PKCS12 -deststoretype JKS -destkeystore 新证书路径\新证书文件名.jks
```

输入上述命令后, 会继续要求用户输入证书的访问密码, 此密码会在 Jmeter 调用证书时使用, 然后提示添加对证书信任选择 y 添加证书信任, 此时会在指定的目录生成符合 Jmeter 格式要求的新证书。

查看新证书的详细信息, 使用如下命令:

```
keytool -list -v -keystore 证书路径\证书文件名.jks
```

2.2.2 Jmeter 配置客户端证书

对客户端证书有两种配置方法: 一种是修改 Jmeter 的 system.properties 文件, 一种是将证书配置信息写入 Beanshell 设置中。比较两种配置方法, 第一种方法证书的配置信息没有写入脚本文件 jmx 中, 当 Jmeter 测试不同应用, 证书需要更改时, 或者脚本发生迁移更换测试环境时都需要重新配置脚本, 第 2 种方法配置信息写入脚本文件, 避免上述需要修改 system.properties 文件的问题, 因而方便脚本迁移与 Jmeter 工具的管理, 因此建议采用第 2 种配置方法。

Beanshell 设置客户端证书的具体方法如下:

在 Jmeter 的【测试计划】的【线程组】中的添加【BeanShell Sampler】采样器, 在采样器的 script 部分添加如下信息:

```
Import org.apache.jmeter.util.SSLManager;
```

```
System.setProperty("javax.net.ssl.keyStore", "证书路径\证书名称.jks");
```

```
System.setProperty("javax.net.ssl.keyStorePassword", "证书访问密码");
```

```
SSLManager.reset();
```

```
SSLManager.getInstance();
```

后两条 SSLManager 函数就是完成将证书加载到 SSL 管理器的操作。

由于 Jmeter 有缓存机制, Jmeter 启动期间只需要运行一次【BeanShell Sampler】采样器即可完成对 TLS1.2 协议的客户端证书配置, 因此将【BeanShell Sampler】采样器可以放置到【仅一次控制器】中。配置完上述信息后重新启动 Jmeter, 加载测试脚本完成客户端证书的配置。对于同一应用系统 https 与 wss 使用一套 TLS1.2 认证证书, 因此不需要为 wss 协议单独配置证书。将客户端证书安装到录制 https 脚本的浏览器的证书目录中。

2.3 安装 PerfMon 插件用于性能测试过程中服务器各项指标监控

PerfMon 的使用主要包含两部分: ServerAgent、PerfMon, ServerAgent 放在监控服务器上; PerfMon 是在 Jmeter 上运行的插件, 用来收集并展示服务器端各项性能指标。在 jmeter-plugins-manage 中下载安装 perfmon 插件, 此部分有成熟的配置方法, 在本文中不做详细描述。

2.4 配置 Jmeter 分布式部署环境

Jmeter 支持多压力机分布式部署, 用于解决性能测试中由于大并发用户数、大数据量传输引起的压力机本身资源及网络带宽的限制。

图 1 是 Jmeter 分布式部署的网络拓扑图, 一台装有 Jmeter 的物理机作为控制台, 其它装有 Jmeter 的物理机作为压力机受控制台调度。修改每台压力机的 Jmeter 的安装目录下的 bin/jmeter.properties 文件的 remote_hosts 属性, 添加本台物理机的 ip 地址与 jmeter 端口号; 修改控制台的 Jmeter 的安装目录下的 bin/jmeter.properties 文件的 remote_hosts 属性, 添加本台物理机与其他压力机的 ip 地址与 jmeter 端口号, 并用逗号隔开。Windows 环境下点击 jmeter-server.bat,

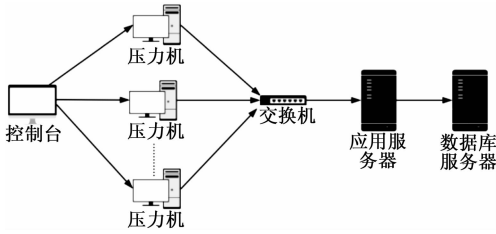


图 1 Jmeter 分布式部署

完成压力机的启动，运行控制台的 jmeter 脚本时可以选择启动指定压力机或者启动全部压力机。

3 编制 https 与 wss 性能测试脚本

3.1 编写 wss 测试脚本

JMeterWebSocketSamplers-0.11.1.jar 是 2.1 节 Jmeter 安装的用于 WebSocket 测试的组件，它支持 wss 协议的主要采样器如下：

- 1) WebSocket Open Connection;
- 2) WebSocket request-response Sampler;
- 3) WebSocket Close.

WebSocket Open Connection: 支持握手阶段，建立 WebSocket 连接，它只提供消息连接的建立而不负责消息传输。WebSocket Close: 客户端主动关闭 WebSocket 请求。一般在 WebSocket request-response Sampler 后使用，用于关闭不需要的 WebSocket 连接，减少服务器连接池的占用。WebSocket request-response Sampler 作为 wss 测试的核心组件，可以完成 wss 连接的创建与消息传输任务。因此本文重点讲述此采样器的使用方法。

WebSocket request-response Sampler 支持 wss 与 ws 两种协议。它的配置页面如图 2 所示，主要由三部分组成，分别是 Connection, Server URL, Data。Connection 用于选择创建连接的方式；Server URL 配置测试的协议种类，服务器的 URL 及访问端口；Data 用于配置的测试数据，连接的超时时间。

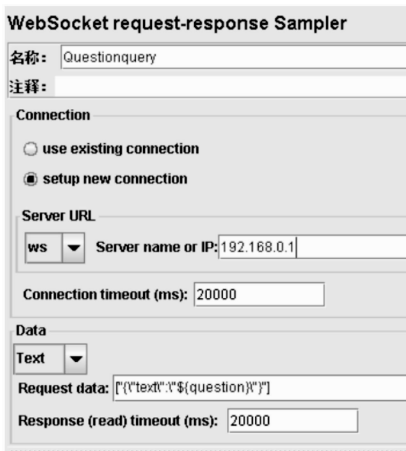


图 2 websocket 请求响应采样器

使用此采样器进行性能测试需要注意的地方是 Connection 部分的连接创建的方式的选择，不同的选择适用于不同

的性能测试场景，下面将简述两种选择方式的区别。第 1 种是使用之前创建完成没有关闭的 WebSocket 连接，例如使用 WebSocket Open Connection 或者 WebSocket request-response Sampler 在之前创建成功的 WebSocket 通信链路，优点是减少连接创建时间请求响应时间更短，节约服务器的连接资源数；第 2 种创建新的 WebSocket 连接，优点是避免使用已有连接进行请求时出现超时失败的情况。分析以上两种连接方式的优点，选择策略如下：当测试的并发请求数较少，对响应时间要求较短的情况，选择第一种使用已创建的连接的方式；当测试的并发请求较多，对响应的成功率要求较高的情况，选择第 2 种方式使用新创建连接。

另外，此采样器 Data 部分配置测试数据时，支持参数变量的引用，可以将测试数据进行参数化，使用不同的测试数据进行性能测试能更好的模拟真实的用户使用场景。

图 3 是 Jmeter 测试结果树控件返回的请求数据，图 4 是 Jmeter 测试结果树控件返回的响应数据。检查请求数据可以判断请求数据参数化调用的正确性；抽取部分响应数据做响应断言，可以判断请求返回的结果的正确性判断，如何选择响应断言数据，将在 4.2 节详述。



图 3 wss 请求体

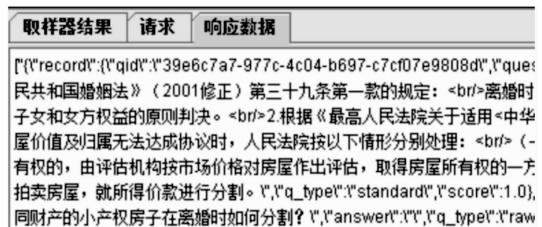


图 4 wss 响应体

3.2 使用 HTTP 代理服务器录制 https 测试脚本

https 脚本可以利用 Jmeter 的【HTTP 代理服务器】，使用浏览器进行录制。此协议的脚本的录制方法较为成熟，此文不予详述。

3.3 在测试脚本中添加监控组件

从监听器中选择【jp@gc - Transactions per Second】、【jp@gc - PerfMon Metrics Collector】与【jp@gc - Response Times Over Time】三个组件作为性能指标监控组件，【jp@gc - Transactions per Second】用于统计每秒的事务数，【jp@gc - PerfMon Metrics Collector】用于监控服务器的 CPU，内存，I/O 使用情况，【jp@gc - Response Times Over Time】用于监控请求的响应时间，同时添加【聚合报告】，用于统计每个事务与请求的响应时间，错误率，吞吐量等详细信

息。在脚本编制阶段可以添加【察看结果树】组件用于查看脚本运行的正确性与辅助优化测试脚本。

由于 Jmeter 自身的监控插件存在局限性, 为了获取更详细的服务器监控数据, Linux 操作系统可以使用 nmon 监控工具, Window 操作系统可以使用系统自带的资源监视器, 并配合 Java 自带的 Jconsole 工具, 与各类数据库监控工具。这些工具的综合使用可以较为全面的监控系统性能测试过程中服务器的各项指标, 查找分析系统的性能瓶颈。

4 性能测试设计

前 3 节主要是针对智能法律问答系统使用的协议 wss 与 https 协议的特殊性, 讲述性能测试工具的选择, 测试工具的配置与脚本的编写。本章节将分析智能法律问答系统的数据与业务特点, 重点讲述针对此类系统的性能测试的设计方法。

4.1 测试数据的选取

分析智能法律问答系统的设计特点, 主要是由自然语言句型识别, 分词处理, 关键词检索几个主要步骤组成^[7]。因此对测试数据的选取也是围绕以上几个步骤, 将作为性能测试输入数据的各类法律问题^[8]按以下规则进行分类:

- 1) 问句中包含系统设定的关键词在系统中的使用频率;
- 2) 问句中包含的系统设定的关键词的数量;
- 3) 问句句型的复杂度, 分词和关键词提取的难易程度;
- 4) 问句所属的业务领域占系统设定的全部业务领域的比例。

根据系统设计的性能指标^[9], 按每个规则将测试数据分为高中低三类, 根据不同的性能测试场景选取不同类别的数据进行测试可以保证数据覆盖的全面性, 真实性。

同时也需要考虑系统已存储的问答对数量对性能测试的影响, 百万级数量的问答对的检索速度会明显与万级数量的问答对的检索速度不同。所以也要将系统存储的问答对的数量级进行等级划分。

4.2 响应结果正确性的验证

不同的法律领域的问题答案显然是不同的, 但相同领域的法律问题回答必然包含相同的法律词汇, 因此可以选取不同法律领域使用较多的法律词汇作为问题答案必须包含的关键词, 作为系统返回的答案的正确性的判断^[10]。将这些关键词用正则表达式组合在一起, 只要系统返回的答案满足正则表达式的设定规则就判断系统返回的结果是正确的。具体操作就是在 wss 请求体加上正则表达式抽取器, 再加上此请求体的响应断言控件, 响应断言判断正则表达式抽取的内容是否与包含此领域问题设定的关键词, 包含则认为系统返回问题答案正确, 不包含则认为系统返回问题答案错误。

4.3 性能测试场景的设定

基于 wss 协议智能法律问答系统的业务特点, 性能测试主要选择压力测试与疲劳测试两种类型^[11]。可选取系统

的主要业务场景包括:

1) 不同规则分类的问题响应时间测试: 使用 4.1 节分类后的测试数据进行压力测试, 用于测试系统在法律领域的广度与问句的复杂度下的性能表现。

2) 不同用户量的测试: 不同数量级的用户数使用同类问题进行压力测试, 测试系统的响应时间与用户承载量。

3) 不同问答对存储数量的压力测试: 向系统提前灌入预估出使用半年, 2 年, 5 年后问答对数量, 分别在系统存储不同数量级的问答对情况下, 测试常规用户量使用下系统的性能表现。

4) 长时间运行稳定性测试: 常规数量用户持续使用问答系统, 在 12 小时, 24 小时, 3 * 24 小时后系统的各项指标的稳定性。

5) 长时间待机可靠性测试: 保持系统应用启动, 在 7 * 24 小时, 31 * 24 小时期间, 利用测试工具间断性的使用问答系统, 验证系统的可用性。

5 结束语

本文针对目前法律领域使用越来越广泛的智能问答系统, 系统地阐述了对此类系统进行性能测试的完整方案。从性能测试的工具选择, 测试工具配置, 脚本编写, 测试数据准备, 测试结果验证及性能场景选取几个方面, 抽取各部分的重点与难点, 提出有效的解决方案, 为进行此类系统性能测试的人员提供必要的参考。

参考文献:

- [1] 周倚文. 面向法律领域的问答系统研究 [D]. 长沙: 湖南大学, 2018: 20-26.
- [2] Tomljanovic J, Pavlic M, Katic M A. Intelligent Question-Answering systems: Review of research [A]. International Convention on Information & Communication Technology, Electronics & Microelectronics [C]. IEEE, 2014: 1478-1483.
- [3] 刘增健. 基于网络搜索的问答系统 [D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [4] 郑强, 徐国胜. Websocket 在服务器推送中的研究 [A]. 中国通信学会学术年会 [C]. 2012: 376-381.
- [5] 付弦. 基于问题库的自动问答智能控制系统研究 [J]. 信息通信, 2016, 158 (2): 4-5.
- [6] 贾文芳, 徐可. 论 Web 应用软件性能测试及测试工具的选择 [J]. 电脑知识与技术, 2011, 07 (20): 4902-4903.
- [7] 曹志娟. 自动问答系统中的问题理解与信息检索研究 [D]. 重庆: 重庆大学, 2005: 18-35.
- [8] 李晓玲, 张茗馨. 百姓工作生活实用法律问答 [M]. 北京: 金盾出版社, 2013.
- [9] 李健. 基于 Web 应用系统的性能测试技术研究 [D]. 西安: 西安电子科技大学, 2010: 34-39.
- [10] 孙胜方, 王昌. 一种人工智能应答系统测试方法及系统: ZL201510346495. 8 [P/OL]. 2015-11-25.
- [11] 赵晗. 基于 WebSocket 协议的 Web 应用性能测试 [D]. 北京: 中国科学院大学, 2014.